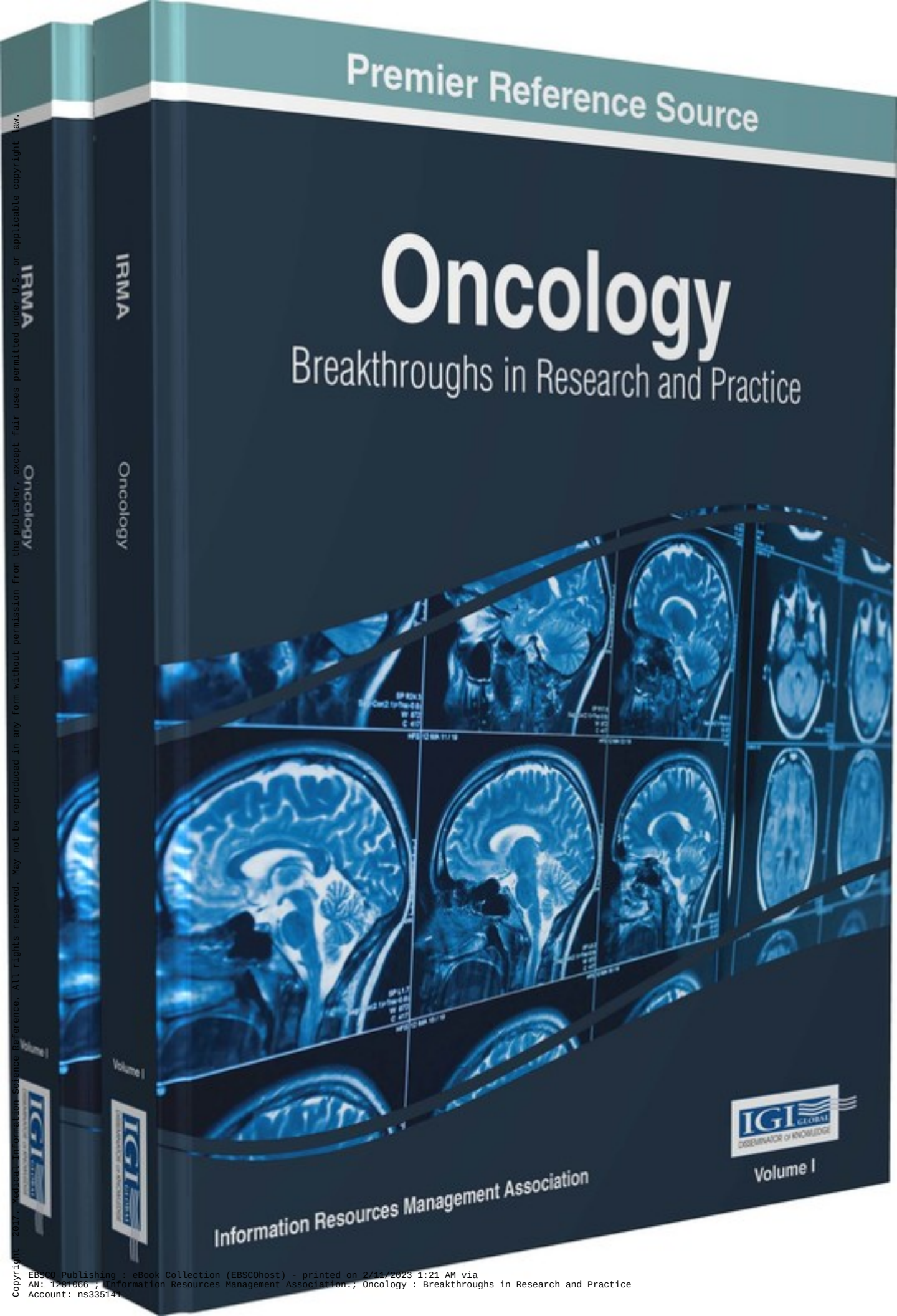


Copyright © 2017, Medical Information Science Reference. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

EBSCO Publishing : eBook Collection (EBSCOhost) - printed on 2/11/2023 1:21 AM via
AN: 1281066 ; Information Resources Management Association.; Oncology : Breakthroughs in Research and Practice
Account: ns335141

Oncology:

Breakthroughs in Research and Practice

Information Resources Management Association
USA

Medical Information Science Reference
An Imprint of IGI Global

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Published in the United States of America by
Medical Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA, USA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

Copyright © 2017 by IGI Global. All rights reserved. No part of this publication may be reproduced, stored or distributed in any form or by any means, electronic or mechanical, including photocopying, without written permission from the publisher. Product or company names used in this set are for identification purposes only. Inclusion of the names of the products or companies does not indicate a claim of ownership by IGI Global of the trademark or registered trademark.

Library of Congress Cataloging-in-Publication Data

Names: Information Resources Management Association, editor.
Title: Oncology : breakthroughs in research and practice / Information Resources Management Association, editor.
Other titles: Oncology (2017)
Description: Hershey, PA : Medical Information Science Reference, [2017] | Chapter 8 is reprint of chapter originally published in Handbook of research on advanced techniques in diagnostic imaging and biomedical applications / [edited by] Themis P. Exarchos, Athanasios Papadopoulos, Dimitrios I. Fotiadis. 2009. | Includes bibliographical references and index.
Identifiers: LCCN 2016013015| ISBN 9781522505495 (hardcover) | ISBN 9781522506386 (ebook)
Subjects: | MESH: Medical Oncology
Classification: LCC RC263 | NLM QZ 200 | DDC 616.99/4--dc23 LC record available at http://lccn.loc.gov/2016013015

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

All work contributed to this book is new, previously-unpublished material. The views expressed in this book are those of the authors, but not necessarily of the publisher.

For electronic access to this publication, please contact: eresources@igi-global.com.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

List of Contributors

Adhikari, Nilanajn / *Jadavpur University, India* 118
Akujuobi, Cajetan M. / *Prairie View A&M University, USA* 352
Al-Mansour, Ebtihal Abdullah / *Al Imam Mohammad Ibn Saud Islamic (IMSIU), Saudi Arabia* 559
Archondakis, Stavros / *401 Army General Hospital, Greece* 267
Asvestas, Pantelis A. / *Technological Educational Institute of Athens, Greece* 689
Baarah, Aladdin / *University of Ottawa, Canada* 187
Bahkali, Salwa / *King Saud Bin Abdulaziz University for Health Sciences, Saudi Arabia* 660
Bawazir, Amin / *King Saud Bin Abdulaziz University for Health Sciences, Saudi Arabia* 660
Bhat, M. Amin / *Sathyabama University, India* 311
Bindra, Jaspreet / *University of Ottawa, Canada* 187
Chamney, Austin / *University of Ottawa, Canada* 187
Chavan, Satishkumar S. / *Don Bosco Institute of Technology, India* 519
Cooper, Malcolm / *Ritsumeikan Asia Pacific University, Japan* 203
Dastmalchi, Siavoush / *Tabriz University of Medical Sciences, Iran* 20,764
Dearden, John C. / *Liverpool John Moores University, UK* 67
Deeb, Omar / *Al-Quds University, Palestine* 408
Deka, Ramesh C. / *Tezpur University, India* 891
Dong, Quoc-Hiep / *University of Medicine and Pharmacy at HCMC, Vietnam* 482
Eid, Mahmoud / *University of Ottawa, Canada* 637,739
Goodarzi, Mohammad / *Katholieke Universiteit Leuven, Belgium* 408
Goyal, Ruchika / *Guru Jambheshwar University of Science and Technology, India* 804
Grover, Aditya / *University of South Florida, USA* 366
Gupta, Neha / *Northeastern University, USA & Osmania University, India* 571
Gupta, Shikhar / *National Institute of Pharmaceutical Education and Research, India* 591
Halder, Amit Kumar / *Jadavpur University, India* 118
Hamzeh-Mivehroud, Maryam / *Tabriz University of Medical Sciences, Iran* 20,764
Hassan, Md. Imtaiyaz / *Jamia Millia Islamia, India* 876
Hieda, Mayumi / *St Luke Clinic, Japan* 203
Hirani, Anjali / *University of South Florida, USA & Virginia Tech-Wake Forest University, USA* 366
Househ, Mowafa / *King Saud Bin Abdulaziz University for Health Sciences, Saudi Arabia* 660
Hua, May / *Columbia University, USA* 1
Isvoran, Adriana / *West University of Timisoara, Romania* 941
Ivascu, Natalia S. / *Weill Cornell Medical College, USA* 728

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Jaffar, Arfan / *Al Imam Mohammad Ibn Saud Islamic (IMSIU), Saudi Arabia* 559
Jain, Sandeep / *Guru Jambheshwar University of Science and Technology, India* 804
Jamil, Uzma / *Bahria University Islamabad, Pakistan & Government College University, Pakistan* 542
Jha, Tarun / *Jadavpur University, India* 118
Kahlon, Amandeep Kaur / *International Centre for Genetic Engineering and Biotechnology, India* 829
Karakitsos, Petros / *University of Athens, Greece* 267
Khalid, Shehzad / *Bahria University Islamabad, Pakistan* 542
Khalifa, Mohamed / *King Faisal Specialist Hospital and Research Center, Saudi Arabia* 660
Kim, Youjeong / *New York Institute of Technology, USA* 676
Kirubakaran, Sivapriya / *Indian Institute of Technology Gandhinagar, India* 342
Kottaridi, Christine / *University of Athens, Greece* 267
Kouloulias, Vasilios / *University of Athens, Greece* 689
Kumar, Akhil / *CSIR-Central Institute of Medicinal and Aromatic Plants, India* 829
Kumar, Ashwani / *Guru Jambheshwar University of Science and Technology, India* 804
Kuziemsky, Craig / *University of Ottawa, Canada* 187
Le, Duy-Phong / *University of Medicine and Pharmacy at HCMC, Vietnam* 482
Le, Minh-Tri / *University of Medicine and Pharmacy at HCMC, Vietnam* 482
Lee, Yong Woo / *Virginia Tech-Wake Forest University, USA* 366
Lone, Imtiyaz H. / *IGNOU Maidan Garhi, India* 311
Markaki, Vasiliki / *National Technical University of Athens, Greece* 689
Marsh, Kate / *Northside Nutrition and Dietetics, Australia* 216
Matsopoulos, George / *National Technical University of Athens, Greece* 689
Meltzer, Ellen C. / *Weill Cornell Medical College, USA* 728
Minovski, Nikola / *National Institute of Chemistry, Slovenia* 434
Mohan, C. Gopi / *Amrita Institute of Medical Sciences and Research Centre, India* 591
Moumtzoglou, Anastasius / *Hellenic Society for Quality & Safety in Healthcare, Greece & P. & A. Kyriakou Children's Hospital, Greece* 960
Nahon-Serfaty, Isaac / *University of Ottawa, Canada* 637,739
Nanda, Anima / *Sathyabama University, India* 311
Naqvi, Ahmad Abu Turab / *Jamia Millia Islamia, India* 876
Nayak, B. K. / *KMC PGS Autonomus Pudicherry, India* 311
Nguyen, Thi-Thanh-Lan / *University of Medicine and Pharmacy at HCMC, Vietnam* 482
Novič, Marjana / *National Institute of Chemistry, Slovenia* 434
Pathak, Yashwant / *University of South Florida, USA* 366
Patnaik, Srinivas / *University of Houston, USA* 382
Peyton, Liam / *University of Ottawa, Canada* 187
Platoni, Kalliopi / *University of Athens, Greece* 689
Pouliakis, Abraham / *University of Athens, Greece* 267,960
Prasad, M. N. Giri / *Jawaharlal Nehru Technological University, India* 302
Rajkhowa, Sanchaita / *Tezpur University, India* 891
Saha, Achintya / *University of Calcutta, India* 118
Satheesha, T. Y. / *Nagarjuna College of Engineering and Technology, India* 302
Sathyanarayana, D. / *Ragiv Gandi Memorial College of Engineering and Technology, India* 302
Saunders, Angela / *Sanitarium Health and Wellbeing, Australia* 216

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Sharma, Ashok / *CSIR-Central Institute of Medicinal and Aromatic Plants, India* 829
Sheikh, Mohamud / *University of New South Wales, Australia* 660
Singh, Anamika / *Maitreyi College, India & University of Delhi, India* 571
Singh, Rajeev / *Indian Council of Medical Research, India* 571
Singh, Vineet / *Shri Govindram Seksaria Institute of Technology and Science, India* 848
Sokouti, Babak / *Tabriz University of Medical Sciences, Iran* 20,764
Spathis, Aris / *University of Athens, Greece* 267
Srivastav, Vijay Kumar / *Shri Govindram Seksaria Institute of Technology and Science, India* 848
Sutariya, Vijaykumar / *University of South Florida, USA* 366
Tabaie, Sheida / *Weill Cornell Medical College, USA* 728
Talbar, Sanjay N. / *S.G.G.S Institute of Engineering and Technology, India* 519
Thai, Khac-Minh / *University of Medicine and Pharmacy at HCMC, Vietnam* 482
Thiruvenkatam, Vijay / *Indian Institute of Technology Gandhinagar, India* 342
Tiwari, Meena / *Shri Govindram Seksaria Institute of Technology and Science, India* 848
Tran, Thanh-Dao / *University of Medicine and Pharmacy at HCMC, Vietnam* 482
Tripathi, Shubhandra / *CSIR-Central Institute of Medicinal and Aromatic Plants, India* 829
Wall, Michael H. / *University of Minnesota, USA* 709
Wang, Jin / *Stony Brook University, USA* 915
Yan, Zhiqiang / *Changchun Institute of Applied Chemistry, China* 915
Zeuschner, Carol / *Sydney Adventist Hospital, Australia* 216

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table of Contents

Preface xiii

Volume I

Section 1
Fundamental Concepts and Design Methodologies

This section serves as a foundation for this exhaustive reference tool by addressing underlying principles essential to the understanding of oncology and provides in-depth coverage of conceptual architecture frameworks to provide the reader with a comprehensive understanding of the emerging developments within the field of oncology. With 9 chapters comprising this foundational section, the reader can learn and chose from a compendium of expert research on the elemental theories underscoring the oncology discipline. Research fundamentals imperative to the understanding of developmental processes within oncology are offered. Chapters found within these pages provide an excellent framework in which to position oncology within the field of information science and technology. From broad examinations to specific discussions on methodology, the research found within this section spans the discipline while offering detailed, specific discussions. From basic designs to abstract development, these chapters serve to expand the reaches of development and design technologies within the oncology community. Insight regarding the critical incorporation of global measures into oncology is addressed, while crucial stumbling blocks of this field are explored.

Chapter 1
Palliative Care 1
May Hua, Columbia University, USA

Chapter 2
An Introduction to the Basic Concepts in QSAR-Aided Drug Design 20
Maryam Hamzeh-Mivehroud, Tabriz University of Medical Sciences, Iran
Babak Sokouti, Tabriz University of Medical Sciences, Iran
Siavoush Dastmalchi, Tabriz University of Medical Sciences, Iran

Chapter 3
The History and Development of Quantitative Structure-Activity Relationships (QSARs) 67
John C. Dearden, Liverpool John Moores University, UK

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 4
Ligand- and Structure-Based Drug Design of Non-Steroidal Aromatase Inhibitors (NSAIs) in Breast Cancer 118
Tarun Jha, Jadavpur University, India
Nilanajn Adhikari, Jadavpur University, India
Amit Kumar Halder, Jadavpur University, India
Achintya Saha, University of Calcutta, India

Chapter 5
A Bounded Health Information Technology System Design Approach to Support Community-Based Care Delivery 187
Liam Peyton, University of Ottawa, Canada
Jaspreet Bindra, University of Ottawa, Canada
Aladdin Baarah, University of Ottawa, Canada
Austin Chamney, University of Ottawa, Canada
Craig Kuziemsky, University of Ottawa, Canada

Chapter 6
The Dark Side of Medical Tourism? End of Life Choice, Human Trafficking, and Organ Transplants 203
Malcolm Cooper, Ritsumeikan Asia Pacific University, Japan
Mayumi Hieda, St Luke Clinic, Japan

Chapter 7
Red Meat and Health: Evidence Regarding Red Meat, Health, and Chronic Disease Risk 216
Kate Marsh, Northside Nutrition and Dietetics, Australia
Angela Saunders, Sanitarium Health and Wellbeing, Australia
Carol Zeuschner, Sydney Adventist Hospital, Australia

Chapter 8
HPV Detection Methods: Towards Personalized Prevention 267
Aris Spathis, University of Athens, Greece
Christine Kottaridi, University of Athens, Greece
Abraham Pouliakis, University of Athens, Greece
Stavros Archondakis, 401 Army General Hospital, Greece
Petros Karakitsos, University of Athens, Greece

Chapter 9
Proposed Threshold Algorithm for Accurate Segmentation for Skin Lesion 302
T. Y. Satheesha, Nagarjuna College of Engineering and Technology, India
D. Sathyanarayana, Ragiv Gandi Memorial College of Engineering and Technology, India
M. N. Giri Prasad, Jawaharlal Nehru Technological University, India

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Section 2
Technologies and Applications

This section presents an extensive coverage of various tools and technologies available in the field of oncology and discusses a variety of applications and opportunities available that can be considered by practitioners in developing viable and effective oncology programs and processes. With 11 chapters, this section offers a broad treatment of some of the many tools and technologies within the oncology field. These chapters enlighten readers about fundamental research on the many tools facilitating the burgeoning field of oncology and review topics from case studies to best practices and ongoing research. Further chapters discuss oncology in a variety of settings. Contributions included in this section provide excellent coverage of today's IT community and how research into oncology is impacting the social fabric of our present-day global village. It is through these rigorously researched chapters that the reader is provided with countless examples of the up-and-coming tools and technologies emerging from the field of oncology.

Chapter 10
Nanotechnology, Metal Nanoparticles, and Biomedical Applications of Nanotechnology 311
M. Amin Bhat, Sathyabama University, India
B. K. Nayak, KMC PGS Autonomus Pudicherry, India
Anima Nanda, Sathyabama University, India
Imtiyaz H. Lone, IGNOU Maidan Garhi, India

Chapter 11
Diverse Applications of Nanotechnology in Biomedicine, Chemistry, and Engineering 342
Sivapriya Kirubakaran, Indian Institute of Technology Gandhinagar, India
Vijay Thiruvenkatam, Indian Institute of Technology Gandhinagar, India

Chapter 12
Nanotechnology Applications in Biomedical Engineering 352
Cajetan M. Akujuobi, Prairie View A&M University, USA

Chapter 13
Nanotechnology for Omics-Based Ocular Drug Delivery 366
Anjali Hirani, University of South Florida, USA & Virginia Tech-Wake Forest University, USA
Aditya Grover, University of South Florida, USA
Yong Woo Lee, Virginia Tech-Wake Forest University, USA
Yashwant Pathak, University of South Florida, USA
Vijaykumar Sutariya, University of South Florida, USA

Chapter 14
Nanomedicine Magic Bullet for Human Cancer 382
Srinivas Patnaik, University of Houston, USA

Chapter 15
QSAR of Antioxidants 408
Omar Deeb, Al-Quds University, Palestine
Mohammad Goodarzi, Katholieke Universiteit Leuven, Belgium

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 16
Integrated in Silico Methods for the Design and Optimization of Novel Drug Candidates: A Case Study on Fluoroquinolones – Mycobacterium tuberculosis DNA Gyrase Inhibitors 434
Nikola Minovski, National Institute of Chemistry, Slovenia
Marjana Novič, National Institute of Chemistry, Slovenia

Volume II

Chapter 17
Computational Approaches for the Discovery of Novel Hepatitis C Virus NS3/4A and NS5B Inhibitors 482
Khac-Minh Thai, University of Medicine and Pharmacy at HCMC, Vietnam
Quoc-Hiep Dong, University of Medicine and Pharmacy at HCMC, Vietnam
Thi-Thanh-Lan Nguyen, University of Medicine and Pharmacy at HCMC, Vietnam
Duy-Phong Le, University of Medicine and Pharmacy at HCMC, Vietnam
Minh-Tri Le, University of Medicine and Pharmacy at HCMC, Vietnam
Thanh-Dao Tran, University of Medicine and Pharmacy at HCMC, Vietnam

Chapter 18
Multimodality Medical Image Fusion Using M-Band Wavelet and Daubechies Complex Wavelet Transform for Radiation Therapy 519
Satishkumar S. Chavan, Don Bosco Institute of Technology, India
Sanjay N. Talbar, S.G.G.S Institute of Engineering and Technology, India

Chapter 19
Analysis of Valuable Techniques and Algorithms Used in Automated Skin Lesion Recognition Systems 542
Uzma Jamil, Bahria University Islamabad, Pakistan & Government College University, Pakistan
Shehzad Khalid, Bahria University Islamabad, Pakistan

Chapter 20
A Study on Automatic Segmentation and Classification of Skin Lesions in Dermoscopic Images 559
Ebtihal Abdullah Al-Mansour, Al Imam Mohammad Ibn Saud Islamic (IMSIU), Saudi Arabia
Arfan Jaffar, Al Imam Mohammad Ibn Saud Islamic (IMSIU), Saudi Arabia

Section 3
Organizational Implications and Critical Issues

This section gives a wide variety of perspectives on oncology and research pertaining to the social and behavioral impact of oncology, for example: the corporate and managerial utilization of information sharing technologies and applications. With 9 chapters, the reader is presented with an in-depth analysis of research into the integration of global oncology as well as implementation of ethical and workflow considerations for all organizations. Chapters introducing this section critically analyze and discuss trends in oncology and crucial questions are addressed and alternatives offered along with theoretical approaches discussed. Also investigating a concern within the field of oncology is research which discusses the effect of user behavior on oncology.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 21
Role of Supercomputers in Bioinformatics 571
Anamika Singh, Maitreyi College, India & University of Delhi, India
Rajeev Singh, Indian Council of Medical Research, India
Neha Gupta, Northeastern University, USA & Osmania University, India

Chapter 22
QSAR Models towards Cholinesterase Inhibitors for the Treatment of Alzheimer's Disease 591
C. Gopi Mohan, Amrita Institute of Medical Sciences and Research Centre, India
Shikhar Gupta, National Institute of Pharmaceutical Education and Research, India

Chapter 23
Ethics, Risk, and Media Intervention: Women's Breast Cancer in Venezuela 637
Mahmoud Eid, University of Ottawa, Canada
Isaac Nahon-Serfaty, University of Ottawa, Canada

Chapter 24
The Role of Primary Health Care in Prevention, Early Detection, and Control of Cancer 660
Salwa Bahkali, King Saud Bin Abdulaziz University for Health Sciences, Saudi Arabia
Mowafa Househ, King Saud Bin Abdulaziz University for Health Sciences, Saudi Arabia
Amin Bawazir, King Saud Bin Abdulaziz University for Health Sciences, Saudi Arabia
Mohamed Khalifa, King Faisal Specialist Hospital and Research Center, Saudi Arabia
Mohamud Sheikh, University of New South Wales, Australia

Chapter 25
My Desired Self, Avatar: The Impact of Avatar Creation on Persuasion 676
Youjeong Kim, New York Institute of Technology, USA

Chapter 26
Isocenter Verification in Radiotherapy Clinical Practice Using Virtual Simulation: An Image Registration Approach 689
George Matsopoulos, National Technical University of Athens, Greece
Pantelis A. Asvestas, Technological Educational Institute of Athens, Greece
Vasiliki Markaki, National Technical University of Athens, Greece
Kalliopi Platoni, University of Athens, Greece
Vasilios Kouloulias, University of Athens, Greece

Chapter 27
Organization and Structure of the Cardiothoracic Intensive Care Unit 709
Michael H. Wall, University of Minnesota, USA

Chapter 28
Ethics 728
Natalia S. Ivascu, Weill Cornell Medical College, USA
Sheida Tabaie, Weill Cornell Medical College, USA
Ellen C. Meltzer, Weill Cornell Medical College, USA

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 29
Risk, Activism, and Empowerment: Women's Breast Cancer in Venezuela 739
Mahmoud Eid, University of Ottawa, Canada
Isaac Nahon-Serfaty, University of Ottawa, Canada

Section 4
Emerging Trends

This section highlights research potential within the field of oncology while exploring uncharted areas of study for the advancement of the discipline. With 9 chapters the readers are introduced to future research directions and topical suggestions for continued debate, centering on the new venues and forums for discussion. In all, this text will serve as a vital resource to practitioners and academics interested in the best practices and applications of the burgeoning field of oncology. The book concludes with a look ahead into the future of the oncology field.

Chapter 30
Molecular Docking at a Glance 764
Maryam Hamzeh-Mivehroud, Tabriz University of Medical Sciences, Iran
Babak Sokouti, Tabriz University of Medical Sciences, Iran
Siavoush Dastmalchi, Tabriz University of Medical Sciences, Iran

Chapter 31
Docking Methodologies and Recent Advances 804
Ashwani Kumar, Guru Jambheshwar University of Science and Technology, India
Ruchika Goyal, Guru Jambheshwar University of Science and Technology, India
Sandeep Jain, Guru Jambheshwar University of Science and Technology, India

Chapter 32
Current Trends in Docking Methodologies 829
Shubhandra Tripathi, CSIR-Central Institute of Medicinal and Aromatic Plants, India
Akhil Kumar, CSIR-Central Institute of Medicinal and Aromatic Plants, India
Amandeep Kaur Kahlon, International Centre for Genetic Engineering and Biotechnology, India
Ashok Sharma, CSIR-Central Institute of Medicinal and Aromatic Plants, India

Chapter 33
Recent Advancements in Docking Methodologies 848
Vijay Kumar Srivastav, Shri Govindram Seksaria Institute of Technology and Science, India
Vineet Singh, Shri Govindram Seksaria Institute of Technology and Science, India
Meena Tiwari, Shri Govindram Seksaria Institute of Technology and Science, India

Chapter 34
Methods for Docking and Drug Designing 876
Ahmad Abu Turab Naqvi, Jamia Millia Islamia, India
Md. Imtaiyaz Hassan, Jamia Millia Islamia, India

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 35
Protein-Ligand Docking Methodologies and Its Application in Drug Discovery 891
Sanchaita Rajkhowa, Tezpur University, India
Ramesh C. Deka, Tezpur University, India

Chapter 36
Scoring Functions of Protein-Ligand Interactions 915
Zhiqiang Yan, Changchun Institute of Applied Chemistry, China
Jin Wang, Stony Brook University, USA

Chapter 37
Online Molecular Docking Resources 941
Adriana Isvoran, West University of Timisoara, Romania

Chapter 38
Mapping Population Health Management Roadmap into Cervical Cancer Screening Programs..... 960
Anastasius Moumtzoglou, Hellenic Society for Quality & Safety in Healthcare, Greece & P. & A. Kyriakou Children's Hospital, Greece
Abraham Pouliakis, ATTIKON University Hospital, Greece

Index xv

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Preface

The constantly changing landscape of oncology makes it challenging for experts and practitioners to stay informed of the field's most up-to-date research. That is why Medical Information Science Reference is pleased to offer this two-volume reference collection that will empower students, researchers, and academicians with a strong understanding of critical issues within oncology by providing both broad and detailed perspectives on cutting-edge theories and developments. This reference is designed to act as a single reference source on conceptual, methodological, and technical, as well as provide insight into emerging trends and future opportunities within the discipline.

Oncology: Breakthroughs in Research and Practice is organized into four sections that provide comprehensive coverage of important topics. The sections are:

1. Fundamental Concepts and Design Methodologies;
2. Technologies and Applications;
3. Organizational Implications and Critical Issues;
4. Emerging Trends;

The following paragraphs provide a summary of what to expect from this invaluable reference tool.

Section 1, "Fundamental Concepts and Design Methodologies," serves as a foundation for this extensive reference tool by addressing crucial theories and presents in-depth coverage of the conceptual design and architecture essential to the understanding of oncology. Introducing the book is *Palliative Care* by May Hua a great foundation laying the groundwork for the basic concepts and theories that will be discussed throughout the rest of the book. Through case studies, this section lays excellent groundwork for later sections that will get into present and future applications for oncology. The section concludes, and leads into the following portion of the book with a nice segue chapter, *Proposed Threshold Algorithm for Accurate Segmentation for Skin Lesion* by T. Y. Satheesha, D. Sathyanarayana, and M. N. Giri Prasad.

Section 2, "Technologies and Applications," presents extensive coverage of the various tools and technologies used along with describing how the broad range of oncology efforts has been utilized and offers insight on and important lessons for their applications and impact. The first chapter *Nanotechnology, Metal Nanoparticles, and Biomedical Applications of Nanotechnology* by M. Amin Bhat, B. K. Nayak, Anima Nanda, and Imtiyaz H. Lone, lays a framework for the types of works that can be found in this section. This section includes the widest range of topics because it describes tools at place in the modeling, planning, and applications along with case studies, research, methodologies, frameworks, architectures, theory, analysis, and guides for implementation of oncology. We conclude the section with *A Study on Automatic Segmentation and Classification of Skin Lesions in Dermoscopic Images* by Ebtihal Abdullah Al-Mansour and Arfan Jaffar a well-rounded transitional chapter into the next section.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Section 3, "Organizational Implications and Critical Issues," includes chapters discussing the organizational and social impact of oncology as well as coverage of academic and research perspectives on oncology tools and applications. The section opens with *Role of Supercomputers in Bioinformatics* by Anamika Singh, Rajeev Singh, and Neha Gupta. This section focuses on how these technologies affect human lives, either through the way they interact with each other, or through how they affect behavioral/workplace situations and also looks into theoretical approaches and offer alternatives to crucial questions on the subject of oncology. The section concludes with *Risk, Activism, and Empowerment: Women's Breast Cancer in Venezuela* by Mahmoud Eid and Isaac Nahon-Serfaty.

Section 4, "Emerging Trends," highlights areas for future research within the field of oncology, opening with *Molecular Docking at a Glance* by Maryam Hamzeh-Mivehroud, Babak Sokouti, and Siavoush Dastmalchi. This section contains chapters that look at what might happen in the coming years that can extend the already staggering amount of applications for oncology. The final chapter of the book looks at an emerging field within oncology, in the excellent contribution, *Mapping Population Health Management Roadmap into Cervical Cancer Screening Programs* by Anastasius Moumtzoglou and Abraham Pouliakis.

Although the primary organization of the contents in this work is based on its four sections, offering a progression of coverage of the important concepts, methodologies, technologies, applications, social issues, and emerging trends, the reader can also identify specific contents by utilizing the extensive indexing system listed at the end. As a comprehensive collection of research on the latest findings related to using technology to providing various services, *Oncology: Breakthroughs in Research and Practice*, provides researchers, administrators and all audiences with a complete understanding of the development of applications and concepts in oncology. Given the vast number of issues concerning usage, failure, success, policies, strategies, and applications of oncology in countries around the world, *Oncology: Breakthroughs in Research and Practice* addresses the demand for a resource that encompasses the most pertinent research in technologies being employed to globally bolster the knowledge and applications of oncology.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Section 1
Fundamental Concepts and Design Methodologies

This section serves as a foundation for this exhaustive reference tool by addressing underlying principles essential to the understanding of oncology and provides in-depth coverage of conceptual architecture frameworks to provide the reader with a comprehensive understanding of the emerging developments within the field of oncology. With 9 chapters comprising this foundational section, the reader can learn and chose from a compendium of expert research on the elemental theories underscoring the oncology discipline. Research fundamentals imperative to the understanding of developmental processes within oncology are offered. Chapters found within these pages provide an excellent framework in which to position oncology within the field of information science and technology. From broad examinations to specific discussions on methodology, the research found within this section spans the discipline while offering detailed, specific discussions. From basic designs to abstract development, these chapters serve to expand the reaches of development and design technologies within the oncology community. Insight regarding the critical incorporation of global measures into oncology is addressed, while crucial stumbling blocks of this field are explored.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 1
Palliative Care

May Hua
Columbia University, USA

ABSTRACT

Palliative care is a specialty of medicine that focuses on improving quality of life for patients with serious illness and their families. As the limitations of intensive care and the long-term sequelae of critical illness continue to be delimited, the role of palliative care for patients that are unable to achieve their original goals of care, as well as for survivors of critical illness, is changing and expanding. The purpose of this chapter is to introduce readers to the specialty of palliative care and its potential benefits for critically ill patients, and to present some of the issues related to the delivery of palliative care in surgical units.

INTRODUCTION

When patients are admitted to the intensive care unit (ICU), the goal of care is most often to rescue and to cure. While this is achievable for the majority of patients, 9% of patients die during their hospitalization (Lilly, 2011). For these patients, the goals of care need to be transitioned, an undertaking that requires the patient, family and all participating members of the care team to agree to these changes. Because of this, the provision of end-of-life care in the ICU is often inadequate, with multiple barriers arising from the patient and family, clinicians and institutional factors (Nelson, 2006b). While patients that die in the ICU have obvious palliative care needs, there is increasing evidence delineating the long-term outcomes of critical illness, with survivors of critical illness being at risk for increased morbidity and mortality and debilitating symptomatology (Herridge; Hofhuis, 2008; Pandharipande, 2013; Wunsch, 2014; Wunsch, 2010). As it becomes clearer that both ICU patients and survivors of critical illness are at risk of having significant symptoms and poor outcomes, it becomes increasingly pressing to determine how to provide the requisite care to meet these needs.

The use of palliative care is one approach that is being advocated to address these concerns. This chapter reviews what palliative care is, evidence for the benefit of palliative care in the ICU setting, models for delivering palliative care in the ICU, barriers to implementing palliative care, particular challenges related to the setting of surgical critical care and discusses future directions for improving the delivery of palliative care and end-of-life care to the surgical patient.

DOI: 10.4018/978-1-5225-0549-5.ch001

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

I. WHAT IS PALLIATIVE CARE?

Traditionally, the goal of medicine has been to cure disease. Hospice and palliative medicine differ in that the goal is to improve quality of life. Palliative care is defined by the World Health Organization as "an approach that improves the quality of life of patients and their families facing the problems associated with life-threatening illness, through the prevention and relief of suffering by means of early identification and impeccable assessment and treatment of pain and other problems, physical, psychosocial and spiritual." As a specialty, palliative care arose out of the hospice movement. Hospice is a model of care for patients with limited life expectancy that focuses on palliation and symptom management, and often requires patients to forego curative therapies. Hospice is most often provided in a patient's home, but can also be provided in hospice inpatient facilities. Unlike hospice, palliative care is appropriate for patients with serious illness at any time in the course of their disease, does not require patients to give up curative treatments, and can be provided in conjunction with other aggressive therapies.

Because the goal of palliative care is to improve quality of life and relieve suffering and stress associated with having a serious illness, the delivery of palliative care may involve many different domains that are utilized based on an individual's specific needs. The use of palliative care addresses various needs for critically ill patients and their families including: 1) symptom management (e.g. pain, nausea, fatigue, thirst, depression, anxiety, existential distress), 2) conflict resolution (conflict between family members, or between the family and the medical team), 3) spiritual support, and 4) complex decision-making regarding goals of care. In the ICU, the role of palliative care has traditionally been as a part of delivering end-of-life care for patients who die in the ICU. However, there is an expanding role for palliative care in the management of patients who may survive their illness but are at risk for having significant symptomatology, such as patients with chronic critical illness.

A. Models for Delivering Palliative Care in the ICU

The two models of delivering palliative care in the ICU are the integrative and consultative models. With the integrative approach, principles of palliative care are woven into routine care, and palliative care is delivered by the primary team, whereas with the consultative approach, palliative care is provided by a specialized palliative care consultation team. These models can co-exist, and reflect the distinction between generalist and specialist palliative care (Quill, 2013). Generalist palliative care is delivered by primary providers, and includes basic pain and symptom management, as well as basic discussions regarding prognosis and goals of care. Specialized palliative care may be required for refractory pain and symptom management, conflict resolution and complex decision-making (Quill, 2013). Each model has its benefits and drawbacks, and which model is optimal likely depends on the characteristics of an individual ICU. Benefits of the integrative model are that there does not need to be additional staffing in order to deliver palliative care, there is less fragmentation of care, and it may decrease conflict associated with having another consulting team. With the consultative model, there is less variability in clinicians' education and their palliative care skill set, and in situations of conflict resolution, the introduction of an impartial party may be helpful. There is a greater body of evidence demonstrating the benefits of specialized palliative care, as a systematic review of communication interventions showed that a consultative approach had better outcomes than an integrative approach (Scheunemann, 2011). However, there is an existing workforce shortage of palliative care physicians (Lupu, 2010), and estimates of the need for

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

palliative care consultation likely outstrips what the current palliative care workforce can provide (Hua, 2014). Consequently, a sustainable model of palliative care delivery in ICUs is likely to be one that is mixed, combining both approaches.

II. EVIDENCE FOR PALLIATIVE CARE IN THE ICU

A. Palliative Care Needs in Critically Ill Patients

In the U.S., admission to an ICU is associated with high in-hospital mortality (Lilly, 2011), and it is estimated that approximately 14% of ICU patients meet screening criteria for palliative care consultation (Hua, 2014). In addition to the large volume of patients with end-of-life care needs, ICU patients may have a substantial symptom burden. Of chronic critically ill patients who were able to report symptoms, 90% endorsed having significant symptoms with uncontrolled pain, psychological distress and dyspnea being most prominent (Nelson, 2004). Cancer patients in the ICU experienced moderate to severe pain, discomfort and anxiety (Delgado-Guay, 2009; Nelson, 2001), and quality of life for patients who die in the ICU is poorer than for those who die at home (Wright, 2010).

Although death frequently occurs during or after an ICU stay, there is also evidence that an ICU admission is a marker of poor long-term prognosis. Survivors of critical illness, particularly those who received mechanical ventilation, have increased 3-year mortality, the bulk of which is concentrated in the first 6 months after hospital discharge (Wunsch, 2010). ICU survivors have also been shown to have significant symptomatology with long-term functional disability, sleep disorders, post-traumatic stress disorder, increased risk of developing a psychological diagnosis and decreased health-related quality of life (Davydow, 2008; Herridge, 2003; Herridge; Hofhuis, 2008; Lee, 2009; Wunsch, 2014). Together, this constellation of symptoms is beginning to carry the label of a post-intensive care syndrome (Harvey, 2012; Mehlhorn, 2014). Furthermore, patients with chronic critical illness who survived to hospital discharge are also known to have a high burden of palliative care needs, with approximately 90% of patients having at least one need (Baldwin, 2013).

Surgical ICU patients, are often perceived to be healthier and to have improved outcomes in comparison to non-surgical ICU patients; of surgical ICU patients, cardiac surgical patients also have lower mortality, with estimates of short-term mortality ranging between 3-4% (Mosenthal, 2012; Nashef, 2012; Shahian, 2012). However, the palliative care needs of patients with surgical critical illness may not differ greatly from their non-surgical counterparts. Mortality remains high for patients with longer ICU stays (Laupland, 2006; Lipsett, 2000), and surgical patients represent a significant portion of the chronic critically ill, a population well-known to be at high-risk for poor outcomes. In a study of the chronic critically ill, 25% of patients transferred to long-term acute care hospitals were surgical, with cardiac surgery patients representing over half of these patients (Kahn, 2013). Furthermore, a study estimating the need for palliative care consultation did not find a significant difference in the percentage of patients meeting screening criteria for palliative care consultation in surgical units in comparison to medical units, but did find that patients in surgical units were significantly less likely to have limitations on treatment placed (Hua, 2014). Thus, the surgical critically ill population may be at higher risk for having unmet palliative care needs.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

B. Benefits of Palliative Care in the ICU

The importance of palliative care in the ICU setting has been accepted and supported by multiple societies, including the American College of Critical Care Medicine and the American Thoracic Society (Lanken, 2008; Truog, 2008). The integration of specialized palliative care in the ICU has been associated with increased rates of formalized advanced directives, increased use of hospice and decreased use of non-beneficial life-sustaining therapies (O'Mahony, 2010). Furthermore, palliative care may also improve costs and resource use. The use of palliative care has been associated with decreased direct costs at the end of life in a generalized hospital setting (Morrison, 2011; Morrison, 2008; Penrod, 2010), and both proactive palliative care consultation and ethics consultations in selected ICU patients were associated with decreased hospital length of stay and decreased use of non-beneficial life-sustaining therapies, without an increase in mortality (Campbell, 2003; Schneiderman, 2003). These outcomes are noteworthy, as non-aggressive care and a family/patient-centered approach to dying are associated with a better quality of death and improved family satisfaction with end-of-life decision making (Gries, 2008; Mularski, 2005). Adequate end-of-life care is highly desired by patients and their families, who may prefer to avoid aggressive and burdensome care in the setting of a poor prognosis (Fried, 2002; Nelson, 2010).

III. BARRIERS TO DELIVERY OF PALLIATIVE CARE

Despite the benefits of palliative care, there are multiple barriers stemming from the culture of the ICU, patients, families, clinicians and institutional factors that ultimately result in the under-utilization of palliative care (Nelson, 2006b).

A. Culture of the ICU

The ICU is the most technologically advanced practice setting in the hospital, where routine care involves the use of machines and medications to monitor and replace organ function for the heart, the lungs, and the kidneys in order to keep patients alive. When a patient is admitted to the ICU, all the stakeholders involved, including the patient, family members and clinicians, often start with the goal of staving off death and "rescuing" the patient. Because of this focus, expectations and hopes are high, and this can lead to a sense of failure, particularly for clinicians, when patients are unable to be "saved" (Nelson, 2006a). When these goals cannot be met, it must first be recognized and accepted by clinicians and families before the goals of care can be changed to providing high-quality end-of-life care. The need for this transition often delays the onset of high-quality end-of-life and palliative care.

B. Patient and Family Members

In a survey of ICU directors, barriers stemming from the patient and family members were rated as the largest barriers to the provision of adequate end-of-life care in the ICU (Nelson, 2006b). These included unrealistic expectations about the patient's prognosis or the effectiveness of therapies, the inability of patients to participate in treatment discussions, a lack of advanced directives and disagreements within

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

families about goals of care (Aslakson, 2012; Nelson, 2006b). Some of the dissonance between families' and clinicians' perceptions may come from a distrust of clinicians' abilities to accurately prognosticate, and family members relying on their own knowledge of a patient to form their own prognosis (Boyd, 2010; Zier, 2012).

C. Clinician and Institutional Factors

Clinicians and institutional factors also present barriers to the delivery of end-of-life and palliative care. In this domain, the most significant barriers reported were that clinicians may not have adequate skills in communication regarding end-of-life care or in techniques for withdrawing or withholding life-sustaining treatment, that there are competing demands for clinicians' time, and that communication with families and other involved clinicians about appropriate goals of care can be deficient (Aslakson, 2012; Friedenberg, 2012; Nelson, 2006b). Barriers that arise from institutional factors include the lack of a palliative care service or palliative care consultants, inadequate support services for family and the lack of a continuity of care due to changes in clinician staffing (Nelson, 2006b).

Many barriers to the adequate delivery of end-of-life care and palliative care have been elicited, but overall, those related to the patient and family (lack of patient involvement in end-of-life care decisions, a lack of advanced directives and unrealistic expectations) have been consistently rated as major barriers, whereas clinician and institutional factors were not reported to be as significant (Friedenberg, 2012; Nelson, 2006b). However, the generalizability of barriers to end-of-life and palliative care may be questionable, as perceived barriers have been shown to differ by level of training, discipline and institution (Friedenberg, 2012). This suggests that to some degree, models and interventions to improve the delivery of palliative care and end-of-life care in the ICU may need to be tailored to specific institutions.

IV. ISSUES SPECIFIC TO CARDIOTHORACIC CRITICAL CARE

A. Surgical Critical Illness

Surgical critical illness differs from non-surgical critical illness in several ways that may impact efforts to initiate end-of-life care or palliative care. First, surgeons are less likely to operate on patients with a do-not-resuscitate order and advanced directives, which may make surgical patients less likely to have them (Bradley, 2010b). Second, unlike many non-surgical patients who are admitted to the ICU for worsening of a chronic condition, surgical patients are often healthier with an acute illness that requires emergent surgery (e.g. trauma), or are undergoing a procedure that is meant to be curative (e.g. aortic valve replacement), or are receiving therapy that is intended to cure their chronic illness (e.g. heart transplantation). Most often, patients, families and surgeons start with a goal of "beating a disease" and this leads to high hopes for recovery (Mosenthal, 2012). This may be particularly pronounced in situations where patients and families have approached surgeons specifically with the intention of "fixing" the problem, and have not yet begun to entertain the possibility that this goal may not be achievable. Consequently, it may be harder to come to terms with a failure of therapy and there needs to be time for all parties to adjust to the idea that the initial goals may not be met.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

B. The Surgical Contract

Surgeons are often perceived as an impediment to the initiation of palliative care and end-of-life care for critically ill patients (Schwarze, 2010). This is explained in part by the complexity of the relationship between the surgeon and a patient, which begins pre-operatively, prior to the onset of any acute critical illness. The idea of a surgical contract, or surgical "buy-in" refers to a process where surgeons, often informally, negotiate a commitment to postoperative care prior to undergoing high-risk surgical procedures (Schwarze, 2010). Surgeons make a commitment to patient when they decide to operate, and they expect a similar commitment from the patient to undergo the post-operative care that they believe is necessary for recovery. As a consequence, surgeons may then feel betrayed or disappointed when approached by families with request to withdraw life-sustaining treatment at a time that they may feel is premature, and they may be unwilling to agree to the withdrawal and may negotiate with families and other caregivers about continuing to pursue aggressive care (Schwarze, 2010). Surgeons also often feel a sense of personal responsibility for adverse postoperative outcomes, and have described feelings of guilt with poor outcomes that may be accentuated when that outcome was unexpected, or when the outcome may involve a surgical error (Schwarze, 2010; Schwarze, 2012). This is compounded by the fact that surgeons are taught to view death as a "failure" (Mosenthal, 2006). While these pre-operative discussions seem to document the risks inherent in the procedure and the potential need for intensive care, the possibility of requiring prolonged life-sustaining treatment is not often discussed, and patients' preferences regarding these treatments are not elicited (Pecanac, 2014). Awareness of the intricacies of this relationship and its existence as a backdrop for post-operative decision-making is crucial for intensivists as they negotiate goals of care between surgeons, patients and families.

C. Lack of Formal Education in Palliative Care for Surgeons

Until recently, education in palliative care was not emphasized as a part of surgical training. This is problematic, as surgeons are unlikely to involve palliative care for symptom management and unfamiliarity with palliative care services has been associated with decreased referral rates to palliative care (Snow, 2009; Tilden, 2009). 84% of surgical residents surveyed did not report receiving any training in palliative care during residency, and informal training in palliative care was also infrequent, where only 50% of surgical residents reported discussing death and dying with a surgical attending (Brasel, 2004; Galante, 2005). Furthermore, surgical residents lacked skills in basic symptom management, as less than half of a group surveyed were proficient in the treatment of cancer-related pain, and competency did not improve as these residents progressed through their training (Brasel, 2004). These deficiencies may correlate to patient outcomes, as a lack of education in end-of-life care has been shown to be associated with decreased willingness to place limitations on care (Forte, 2012). Also, in a study of palliative care in the ICU, nurse ratings' of the quality of death were lowest for patients cared for by surgical intensivists, and patients cared for by surgical intensivists were significantly less likely to have indicators of palliative care documented in their medical charts (Kross, 2014). As surgeons may be unaware of what palliative care is and its benefits, it is not surprising that surgeons may be more likely to only involve palliative care at the end-of-life (Tilden, 2009).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

D. Conflict between Surgeons and ICU Clinicians

With surgical patients, there is always another highly-invested clinician involved whose relationship with the patient exists prior to the onset of critical illness, and this creates an extra layer of potential conflict that surrounds decision-making. Surgeons are perceived by ICU clinicians as an impediment to end-of-life and palliative care (Schwarze, 2010), and surgeons perceive conflict frequently with ICU clinicians, with approximately 40% of surveyed surgeons reporting that they experienced conflict with intensivists or nurses during the care of one of their patients (Paul Olson, 2013). Much of this conflict may revolve around prognostication, which is often at the center of discussion regarding appropriate goals of care. In a cross-sectional survey, the majority of surgeons were satisfied with their own communication regarding prognosis to patients and families, but less than a quarter of ICU physicians and less than 5% of ICU nurses reported being satisfied with the surgeon's communication of prognosis (Aslakson, 2010). Furthermore, ICU physicians reported having fewer opportunities to communicate with surgeons regarding their concerns about prognosis, and did not report feeling that their comments were valued (Aslakson, 2010). This additional layer of conflict complicates the process of addressing goals of care, and consequently, may hinder the initiation of palliative care and end-of-life care.

E. Staffing Model

Many surgical patients are cared for in open units, with surgeons retaining primary responsibility for the patient's care. Organizational aspects of this particular staffing model present unique challenges to decision-making and palliative care. First, it may be ambiguous which clinician, the surgeon or the intensivist, is responsible for leading discussions regarding goals of care and for coordinating care. Also, the presence of the surgical team in the ICU may vary throughout the day, leading to fragmentation of communication and the overall care plan (Mosenthal, 2012). Lastly, with regards to quality improvement, it may be more difficult to effect cultural changes or implement interventions to improve palliative care and end-of-life care in open units in comparison to closed units (Mosenthal, 2006).

F. Mechanical Circulatory Support Devices

Although the ICU has always represented a forefront of medicine where technology is employed to provide life support and prolong life, the advent of mechanical circulatory support devices (MCSD) such as the ventricular assist device (VAD) and extracorporeal membrane oxygenation (ECMO) adds a new layer of complexity to the provision of palliative care and end-of-life care in a cardiothoracic ICU. These patients are at increased risk for having palliative care needs, and being considered as a candidate for a VAD or having had a cardiac arrest have been suggested as triggers for palliative care consultation (Norton, 2007; Weissman, 2011). Patients requiring mechanical circulatory support are at high risk of mortality, may suffer devastating complications (e.g. stroke, limb ischemia, life-threatening bleeding), and often require complex decision-making regarding goals of care. As an example, consider a patient who receives a VAD as a bridge-to-transplant and then suffers a large stroke that precludes transplantation. While the initial goal of VAD therapy was to support the patient until transplantation, this goal is no longer applicable, and new goals of care need to be defined. In the case of ECMO, goals of care also may rapidly change, particularly if the patient's condition has not improved and continuing care becomes futile. The role of palliative care consultations for ECMO patients has yet to be well-defined

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

(Abrams, 2014), but may be particularly helpful in situations with high prognostic uncertainty and medical complexity (Doorenbos, 2013).

The use of MCSD also may complicate withdrawal of life-sustaining therapy at the end of life. Families may have difficulty with the concept of discontinuing device support, and clinicians may also be reluctant due to ethical and legal concerns (Swetz, 2011). There are also technical issues with discontinuing MCSD, requiring staff to be familiar with each specific device and the proper method of discontinuing support to avoid causing the device to alarm, which may cause distress to families. Consequently, having a protocol in place for device withdrawal has been advocated (Brush, 2010; MacIver, 2005).

V. METHODS TO IMPROVE DELIVERY OF PALLIATIVE CARE IN THE CARDIOTHORACIC ICU

General recommendations for developing and integrating palliative care in a surgical ICU setting have been proposed with three steps outlined: 1) form an interdisciplinary workgroup of key stakeholders to lead the effort to implement palliative care, 2) choose a model for implementing palliative care and 3) implement screening triggers for palliative care consultation (Mosenthal, 2012). To start, the workgroup should be comprised of representatives from the leadership of all involved parties, including intensivists, surgeons, cardiologists, critical care nurses, palliative care specialists, and if possible, hospital leadership. For the second step, the needs and resources for palliative care that are particular to the institution should be assessed to determine how much each model (consultative, integrative, or mixed) should be employed. For the integrative model, one particular approach has been described where a palliative care intervention is implemented for all patients, irrespective of prognosis (Table 1) (Mosenthal, 2006). This has the advantage of not being concerned with prognostication, which is problematic for clinicians and may be difficult for families to accept (Meadow, 2011; Zier, 2012).

In order to facilitate use of the consultative model, choosing appropriate screening triggers for palliative care consultation is recommended. It is estimated that approximately 14% of all ICU patients meet published triggers for palliative care consultation (Hua, 2014). However, whether or not this strategy is feasible in the setting of a cardiothoracic ICU is not known. Triggers have been developed specifically for use in a surgical ICU setting, but when implemented, they failed to increase the number of palliative care consultations (Bradley, 2010a). This may have been secondary to the triggers occurring infrequently or surgeons being unwilling to involve palliative care consultants. Triggers may be more useful in a cardiothoracic ICU if they target specific populations, (e.g. transplant patients, VAD patients, ECMO patients), or if referral is mandated upon meeting a trigger (Mosenthal, 2012).

Changing culture is also a vital component of improving the delivery of palliative care in a cardiothoracic ICU. Surgeons and intensivists may require education about the role of palliative care to understand that palliative care can be provided in conjunction with aggressive therapies. In particular for surgeons, this may be necessary to alleviate fears that palliative care is simply "death care," or that they will lose control of decision-making regarding goals of care. However, for this type of change to occur, the provision of high-quality palliative care must be viewed as being equally important to "rescuing" patients and emphasized as an important goal by the leadership of all key stakeholders.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 1. Checklist of a palliative care intervention in a surgical/trauma intensive care unit(Mosenthal, 2006)*

Intervention	Goal
1. Bereavement and psychosocial support for patients and families within 24 hours of admission to the surgical ICU	To establish a supportive relationship with families early on, which can aid future discussions and decision-making if needed
2. Interdisciplinary palliative care assessment within 24 hours of admission - Assessment of pain and symptoms - Assessment of psychosocial, spiritual and bereavement needs of family - Identify surrogate decision maker - Solicit and clarify advanced directives - Assessment of likely outcome and prognosis (discharge from ICU, discharge from hospital) from the current illness with regards to mortality and quality of life	Focus on symptom management and relief, identify conflicts related to goals of care early
3. Family meeting with physician and nurse within 72 hours of admission to the surgical ICU	To communicate information to families regarding outcomes, address goals of care and model shared decision-making
4. Comprehensive palliative care plan by 72 hours	To document a palliative care plan for all patients who remain in the ICU for ≥ 3 days, including pain and symptom management, goals of care, family support and ongoing communication, irrespective of prognosis To institute limitations in care if appropriate
5. Implementation of Palliative Care Standing Order Set for patients who are identified as imminently dying	To facilitate a standardized delivery of palliative care measures for dying patients
6. Integration of palliative care performance into morbidity and mortality conferences and peer review	To raise awareness of the importance of high-quality palliative care by making the quality of palliative care delivery and the quality of death a formal part of quality improvement

*Adapted from Mosenthal and Murphy.

VI. FUTURE DIRECTIONS

Research in several particular domains is crucial to improving palliative care and end-of-life care for critically ill surgical patients. Different models of delivering palliative care and methods for successful implementation in a cardiothoracic ICU setting require further study, as few successful models have been reported. Approaches that account for the differences in trajectories of illness for cardiac surgical patients, organizational differences in the ICU and the reality of the surgical contract need to be developed and examined to determine their efficacy. Another area that merits investigation is how to improve education in generalist palliative care for both surgeons and intensivists. As consultants are unlikely to be able to fully meet the palliative care needs of cardiothoracic ICU patients, enhancing the palliative care skill set of “front-line” clinicians will not only improve palliative care for patients, but increase the involvement of specialist palliative care when appropriate. Also, there is a need to standardize the manner in which the quality and effectiveness of palliative care in ICUs is measured. If palliative care in the ICU is to be improved, outcomes and measurements that can reliably differentiate high-quality from low-quality palliative care need to be developed and employed consistently. Lastly, given the increasing use of MCSD, this group of patients at high-risk of having palliative care needs will continue to pose a challenge for ICU clinicians. Identifying the palliative care needs of this particular patient population is a necessary first step towards understanding how these needs may be met.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

VII. CONCLUSION

Palliative care in the ICU is appropriate for patients with significant symptomatology, psychological or spiritual distress or conflict regarding goals of care. Adequate delivery of palliative care for all patients is likely to involve a combination of basic palliative care that is delivered by ICU clinicians, and specialized palliative care that is delivered by palliative care consultants. In cardiothoracic surgical units, particular challenges exist to the delivery of palliative care that relate to the surgical ethos of rescuing the patient, the involvement of another clinician that is deeply invested in the patient's outcome and the perceptions of cardiac surgical patients as being relatively healthy and more robust by both families and clinicians. In particular, the population of patients requiring mechanical circulatory support represents a group of patients that are potentially high-risk for having palliative care needs. Efforts to improve palliative care delivery to cardiac surgical critically ill patients should focus on the continuing education of surgeons and intensivists as to what palliative care is and its potential benefits, determining which model of palliative care delivery is most appropriate for cardiac surgical patients, and studying the implementation of these models, and defining the role of palliative care for particular high-risk populations.

REFERENCES

Abrams, D. C., Prager, K., Blinderman, C. D., Burkart, K. M., & Brodie, D. (2014). Ethical dilemmas encountered with the use of extracorporeal membrane oxygenation in adults. *Chest*, *145*(4), 876–882. doi:10.1378/chest.13-1138 PMID:24687709

Aslakson, R. A., Wyskiel, R., Shaeffer, D., Zyra, M., Ahuja, N., Nelson, J. E., & Pronovost, P. J. (2010). Surgical intensive care unit clinician estimates of the adequacy of communication regarding patient prognosis. *Critical Care (London, England)*, *14*(6), R218. doi:10.1186/cc9346 PMID:21114837

Aslakson, R. A., Wyskiel, R., Thornton, I., Copley, C., Shaffer, D., Zyra, M., & Pronovost, P. J. et al. (2012). Nurse-perceived barriers to effective communication regarding prognosis and optimal end-of-life care for surgical ICU patients: A qualitative exploration. *Journal of Palliative Medicine*, *15*(8), 910–915. doi:10.1089/jpm.2011.0481 PMID:22676315

Baldwin, M. R., Wunsch, H., Reyfman, P. A., Narain, W. R., Blinderman, C. D., Schluger, N. W., & Bach, P. (2013). *High Burden of Palliative Needs among Older ICU Survivors Transferred to Post-Acute Care Facilities: A Single Center Study*. Ann Am Thorac Soc. doi:10.1513/AnnalsATS.201303-039OC

Boyd, E. A., Lo, B., Evans, L. R., Malvar, G., Apatira, L., Luce, J. M., & White, D. B. (2010). "It's not just what the doctor tells me:" factors that influence surrogate decision-makers' perceptions of prognosis. *Critical Care Medicine*, *38*(5), 1270–1275. doi:10.1097/CCM.0b013e3181d8a217 PMID:20228686

Bradley, C., Weaver, J., & Brasel, K. (2010a). Addressing access to palliative care services in the surgical intensive care unit. *Surgery*, *147*(6), 871–877. doi:10.1016/j.surg.2009.11.005 PMID:20097397

Bradley, C. T., Brasel, K. J., & Schwarze, M. L. (2010b). Physician attitudes regarding advance directives for high-risk surgical patients: A qualitative analysis. *Surgery*, *148*(2), 209–216. doi:10.1016/j.surg.2010.05.020 PMID:20580048

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Brasel, K. J., & Weissman, D. E. (2004). Palliative care education for surgeons. *Journal of the American College of Surgeons*, *199*(3), 495–499. doi:10.1016/j.jamcollsurg.2004.06.004 PMID:15325621

Brush, S., Budge, D., Alharethi, R., McCormick, A. J., MacPherson, J. E., Reid, B. B., & Kfoury, A. G. et al. (2010). End-of-life decision making and implementation in recipients of a destination left ventricular assist device. *The Journal of Heart and Lung Transplantation*, *29*(12), 1337–1341. doi:10.1016/j.healun.2010.07.001 PMID:20817564

Campbell, M. L., & Guzman, J. A. (2003). Impact of a proactive approach to improve end-of-life care in a medical ICU. *Chest*, *123*(1), 266–271. doi:10.1378/chest.123.1.266 PMID:12527629

Davydow, D. S., Gifford, J. M., Desai, S. V., Needham, D. M., & Bienvenu, O. J. (2008). Posttraumatic stress disorder in general intensive care unit survivors: A systematic review. *General Hospital Psychiatry*, *30*(5), 421–434. doi:10.1016/j.genhosppsych.2008.05.006 PMID:18774425

Delgado-Guay, M. O., Parsons, H. A., Li, Z., Palmer, L. J., & Bruera, E. (2009). Symptom distress, interventions, and outcomes of intensive care unit cancer patients referred to a palliative care consult team. *Cancer*, *115*(2), 437–445. doi:10.1002/cncr.24017 PMID:19107768

Doorenbos, A. Z., Starks, H., Bourget, E., McMullan, D. M., Lewis-Newby, M., Rue, T. C., & Wilfond, B. S. et al. (2013). Examining palliative care team involvement in automatic consultations for children on extracorporeal life support in the pediatric intensive care unit. *Journal of Palliative Medicine*, *16*(5), 492–495. doi:10.1089/jpm.2012.0536 PMID:23540309

Forte, D. N., Vincent, J. L., Velasco, I. T., & Park, M. (2012). Association between education in EOL care and variability in EOL practice: A survey of ICU physicians. *Intensive Care Medicine*, *38*(3), 404–412. doi:10.1007/s00134-011-2400-4 PMID:22222566

Fried, T. R., Bradley, E. H., Towle, V. R., & Allore, H. (2002). Understanding the treatment preferences of seriously ill patients. *The New England Journal of Medicine*, *346*(14), 1061–1066. doi:10.1056/NEJMsa012528 PMID:11932474

Friedenberg, A. S., Levy, M. M., Ross, S., & Evans, L. E. (2012). Barriers to end-of-life care in the intensive care unit: Perceptions vary by level of training, discipline, and institution. *Journal of Palliative Medicine*, *15*(4), 404–411. doi:10.1089/jpm.2011.0261 PMID:22468773

Galante, J. M., Bowles, T. L., Khatri, V. P., Schneider, P. D., Goodnight, J. E. Jr, & Bold, R. J. (2005). Experience and attitudes of surgeons toward palliation in cancer. *Archives of Surgery*, *140*(9), 873–878, discussion 878–880. doi:10.1001/archsurg.140.9.873 PMID:16172296

Gries, C. J., Curtis, J. R., Wall, R. J., & Engelberg, R. A. (2008). Family member satisfaction with end-of-life decision making in the ICU. *Chest, 133*(3), 704-712. doi: chest.07-1773 [pii] 10.1378/chest.07-1773

Harvey, M. A. (2012). The truth about consequences--post-intensive care syndrome in intensive care unit survivors and their families. *Critical Care Medicine*, *40*(8), 2506–2507. doi:10.1097/CCM.0b013e318258e943 PMID:22809925

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Herridge, M. S., Cheung, A. M., Tansey, C. M., Matte-Martyn, A., Diaz-Granados, N., Al-Saidi, F., & Slutsky, A. S. et al. (2003). One-year outcomes in survivors of the acute respiratory distress syndrome. *The New England Journal of Medicine*, *348*(8), 683–693. doi:10.1056/NEJMoa022450 PMID:12594312

Herridge, M. S., Tansey, C. M., Matte, A., Tomlinson, G., Diaz-Granados, N., Cooper, A., & Cheung, A. M. et al. (2011, April07). Cheung, A. M. Functional disability 5 years after acute respiratory distress syndrome. *The New England Journal of Medicine*, *364*(14), 1293–1304. doi:10.1056/NEJMoa1011802 PMID:21470008

Hofhuis, J. G., Spronk, P. E., van Stel, H. F., Schrijvers, G. J., Rommes, J. H., & Bakker, J. (2008). The impact of critical illness on perceived health-related quality of life during ICU treatment, hospital stay, and after hospital discharge: A long-term follow-up study. *Chest*, *133*(2), 377–385. doi:10.1378/chest.07-1217 PMID:17925419

Hua, M. S., Li, G., Blinderman, C. D., & Wunsch, H. (2014). Estimates of the Need for Palliative Care Consultation across United States Intensive Care Units Using a Trigger-based Model. *American Journal of Respiratory and Critical Care Medicine*, *189*(4), 428–436. doi:10.1164/rccm.201307-1229OC PMID:24261961

Kahn, J. M., Werner, R. M., David, G., Ten Have, T. R., Benson, N. M., & Asch, D. A. (2013). Effectiveness of long-term acute care hospitalization in elderly patients with chronic critical illness. *Medical Care*, *51*(1), 4–10. doi:10.1097/MLR.0b013e31826528a7 PMID:22874500

Kross, E. K., Engelberg, R. A., Downey, L., Cuschieri, J., Hallman, M. R., Longstreth, W. T. Jr, & Curtis, J. R. et al. (2014). Differences in end-of-life care in the ICU across patients cared for by medicine, surgery, neurology, and neurosurgery physicians. *Chest*, *145*(2), 313–321. doi:10.1378/chest.13-1351 PMID:24114410

Lanken, P. N., Terry, P. B., Delisser, H. M., Fahy, B. F., Hansen-Flaschen, J., & Heffner, J. E. et al. Force, A. T. S. E.-o.-L. C. T. (2008). An official American Thoracic Society clinical policy statement: Palliative care for patients with respiratory diseases and critical illnesses. *American Journal of Respiratory and Critical Care Medicine*, *177*(8), 912–927. doi:10.1164/rccm.200605-587ST PMID:18390964

Laupland, K. B., Kirkpatrick, A. W., Kortbeek, J. B., & Zuege, D. J. (2006). Long-term mortality outcome associated with prolonged admission to the ICU. *Chest*, *129*(4), 954–959. doi:10.1378/chest.129.4.954 PMID:16608944

Lee, C. M., Herridge, M. S., Gabor, J. Y., Tansey, C. M., Matte, A., & Hanly, P. J. (2009). Chronic sleep disorders in survivors of the acute respiratory distress syndrome. *Intensive Care Medicine*, *35*(2), 314–320. doi:10.1007/s00134-008-1277-3 PMID:18802684

Lilly, C. M., Zuckerman, I. H., Badawi, O., & Riker, R. R. (2011). Benchmark data from more than 240,000 adults that reflect the current practice of critical care in the United States. *Chest*, *140*(5), 1232–1242. doi:10.1378/chest.11-0718 PMID:21868469

Lipsett, P. A., Swoboda, S. M., Dickerson, J., Ylitalo, M., Gordon, T., Breslow, M., & Rosenfeld, B. et al. (2000). Survival and functional outcome after prolonged intensive care unit stay. *Annals of Surgery*, *231*(2), 262–268. doi:10.1097/00000658-200002000-00016 PMID:10674619

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Lupu, D. (2010). Estimate of current hospice and palliative medicine physician workforce shortage. *Journal of Pain and Symptom Management*, *40*(6), 899–911. doi:10.1016/j.jpainsymman.2010.07.004 PMID:21145468

MacIver, J., & Ross, H. J. (2005). Withdrawal of ventricular assist device support. *Journal of Palliative Care*, *21*(3), 151–156. PMID:16334969

Meadow, W., Pohlman, A., Frain, L., Ren, Y., Kress, J. P., Teuteberg, W., & Hall, J. (2011). Power and limitations of daily prognostications of death in the medical intensive care unit. *Critical Care Medicine*, *39*(3), 474–479. doi:10.1097/CCM.0b013e318205df9b PMID:21150582

Mehlhorn, J., Freytag, A., Schmidt, K., Brunkhorst, F. M., Graf, J., Troitzsch, U., & Gensichen, J. et al. (2014). Rehabilitation Interventions for Postintensive Care Syndrome: A Systematic Review. *Critical Care Medicine*, *42*(5), 1263–1271. doi:10.1097/CCM.0000000000000148 PMID:24413580

Morrison, R. S., Dietrich, J., Ladwig, S., Quill, T., Sacco, J., Tangeman, J., & Meier, D. E. (2011). Palliative care consultation teams cut hospital costs for Medicaid beneficiaries. *Health Affairs*, *30*(3), 454–463. doi:10.1377/hlthaff.2010.0929 PMID:21383364

Morrison, R. S., Penrod, J. D., Cassel, J. B., Caust-Ellenbogen, M., Litke, A., & Spragens, L. et al.. (2008). Cost savings associated with US hospital palliative care consultation programs. *Archives of Internal Medicine*, *168*(16), 1783–1790. doi:10.1001/archinte.168.16.1783 PMID:18779466

Mosenthal, A. C., & Murphy, P. A. (2006). Interdisciplinary model for palliative care in the trauma and surgical intensive care unit: Robert Wood Johnson Foundation Demonstration Project for Improving Palliative Care in the Intensive Care Unit. *Critical Care Medicine*, *34*(11Suppl), S399–S403. doi:10.1097/01.CCM.0000237044.79166.E1 PMID:17057605

Mosenthal, A. C., Weissman, D. E., Curtis, J. R., Hays, R. M., Lustbader, D. R., Mulkerin, C., & Nelson, J. E. et al. (2012). Integrating palliative care in the surgical and trauma intensive care unit: A report from the Improving Palliative Care in the Intensive Care Unit (IPAL-ICU) Project Advisory Board and the Center to Advance Palliative Care. *Critical Care Medicine*, *40*(4), 1199–1206. doi:10.1097/CCM.0b013e31823bc8e7 PMID:22080644

Mularski, R. A., Heine, C. E., Osborne, M. L., Ganzini, L., & Curtis, J. R. (2005). Quality of dying in the ICU: Ratings by family members. *Chest*, *128*(1), 280–287. doi:10.1378/chest.128.1.280 PMID:16002947

Nashef, S. A., Roques, F., Sharples, L. D., Nilsson, J., Smith, C., Goldstone, A. R., & Lockowandt, U. (2012). EuroSCORE II. *European Journal of Cardio-Thoracic Surgery, 41*(4), 734-744; discussion 744-735. doi: 10.1093/ejcts/ezs043

Nelson, J. E. (2006a). Identifying and overcoming the barriers to high-quality palliative care in the intensive care unit. *Critical Care Medicine*, *34*(11Suppl), S324–S331. doi:10.1097/01.CCM.0000237249.39179.B1 PMID:17057594

Nelson, J. E., Angus, D. C., Weissfeld, L. A., Puntillo, K. A., Danis, M., & Deal, D. et al. Critical Care Peer Workgroup of the Promoting Excellence in End-of-Life Care, P. (2006b). End-of-life care for the critically ill: A national intensive care unit survey. *Critical Care Medicine*, *34*(10), 2547–2553. doi:10.1097/01.CCM.0000239233.63425.1D PMID:16932230

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Nelson, J. E., Meier, D. E., Litke, A., Natale, D. A., Siegel, R. E., & Morrison, R. S. (2004). The symptom burden of chronic critical illness. *Crit Care Med, 32*(7), 1527-1534. doi: 00003246-200407000-00010 [pii]

Nelson, J. E., Meier, D. E., Oei, E. J., Nierman, D. M., Senzel, R. S., Manfredi, P. L., & Morrison, R. S. et al. (2001). Self-reported symptom experience of critically ill cancer patients receiving intensive care. *Critical Care Medicine*, *29*(2), 277–282. doi:10.1097/00003246-200102000-00010 PMID:11246306

Nelson, J. E., Puntillo, K. A., Pronovost, P. J., Walker, A. S., McAdam, J. L., Ilaoa, D., & Penrod, J. (2010). In their own words: Patients and families define high-quality palliative care in the intensive care unit. *Critical Care Medicine*, *38*(3), 808–818. doi:10.1097/CCM.0b013e3181c5887c PMID:20198726

Norton, S. A., Hogan, L. A., Holloway, R. G., Temkin-Greener, H., Buckley, M. J., & Quill, T. E. (2007). Proactive palliative care in the medical intensive care unit: Effects on length of stay for selected high-risk patients. *Critical Care Medicine*, *35*(6), 1530–1535. doi:10.1097/01.CCM.0000266533.06543.0C PMID:17452930

O'Mahony, S., McHenry, J., Blank, A. E., Snow, D., Eti Karakas, S., Santoro, G., & Kvetan, V. et al. (2010). Preliminary report of the integration of a palliative care team into an intensive care unit. *Palliative Medicine*, *24*(2), 154–165. doi:10.1177/0269216309346540 PMID:19825893

Pandharipande, P. P., Girard, T. D., Jackson, J. C., Morandi, A., Thompson, J. L., Pun, B. T., & Ely, E. W. et al. (2013). Long-term cognitive impairment after critical illness. *The New England Journal of Medicine*, *369*(14), 1306–1316. doi:10.1056/NEJMoa1301372 PMID:24088092

Paul Olson, T. J., Brasel, K. J., Redmann, A. J., Alexander, G. C., & Schwarze, M. L. (2013). Surgeon-reported conflict with intensivists about postoperative goals of care. *JAMA Surgery*, *148*(1), 29–35. doi:10.1001/jamasurgery.2013.403 PMID:23324837

Pecanac, K. E., Kehler, J. M., Brasel, K. J., Cooper, Z., Steffens, N. M., McKneally, M. F., & Schwarze, M. L. (2014). It's big surgery: Preoperative expressions of risk, responsibility, and commitment to treatment after high-risk operations. *Annals of Surgery*, *259*(3), 458–463. doi:10.1097/SLA.0000000000000314 PMID:24253139

Penrod, J. D., Deb, P., Dellenbaugh, C., Burgess, J. F. Jr, Zhu, C. W., Christiansen, C. L., & Morrison, R. S. et al. (2010). Hospital-based palliative care consultation: Effects on hospital cost. *Journal of Palliative Medicine*, *13*(8), 973–979. doi:10.1089/jpm.2010.0038 PMID:20642361

Quill, T. E., & Abernethy, A. P. (2013). Generalist plus Specialist Palliative Care - Creating a More Sustainable Model. *The New England Journal of Medicine*, *368*(13), 1173–1175. doi:10.1056/NEJMp1215620 PMID:23465068

Scheunemann, L. P., McDevitt, M., Carson, S. S., & Hanson, L. C. (2011). Randomized, controlled trials of interventions to improve communication in intensive care: A systematic review. *Chest*, *139*(3), 543–554. doi:10.1378/chest.10-0595 PMID:21106660

Schneiderman, L. J., Gilmer, T., Teetzel, H. D., Dugan, D. O., Blustein, J., Cranford, R., & Young, E. W. et al. (2003). Effect of ethics consultations on nonbeneficial life-sustaining treatments in the intensive care setting: A randomized controlled trial. *Journal of the American Medical Association*, *290*(9), 1166–1172. doi:10.1001/jama.290.9.1166 PMID:12952998

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Schwarze, M. L., Bradley, C. T., & Brasel, K. J. (2010). Surgical "buy-in": The contractual relationship between surgeons and patients that influences decisions regarding life-supporting therapy. *Critical Care Medicine*, *38*(3), 843–848. doi:10.1097/CCM.0b013e3181cc466b PMID:20048678

Schwarze, M. L., Redmann, A. J., Brasel, K. J., & Alexander, G. C. (2012). The role of surgeon error in withdrawal of postoperative life support. *Annals of Surgery*, *256*(1), 10–15. doi:10.1097/SLA.0b013e3182580de5 PMID:22584696

Shahian, D. M., O'Brien, S. M., Sheng, S., Grover, F. L., Mayer, J. E., Jacobs, J. P., & Edwards, F. H. et al. (2012). Predictors of long-term survival after coronary artery bypass grafting surgery: Results from the Society of Thoracic Surgeons Adult Cardiac Surgery Database (the ASCERT study). *Circulation*, *125*(12), 1491–1500. doi:10.1161/CIRCULATIONAHA.111.066902 PMID:22361330

Snow, C. E., Varela, B. R., Pardi, D. A., Adelman, R. D., Said, S., & Reid, M. C. (2009). Identifying factors affecting utilization of an inpatient palliative care service: A physician survey. *Journal of Palliative Medicine*, *12*(3), 231–237. doi:10.1089/jpm.2008.0235 PMID:19254200

Swetz, K. M., Ottenberg, A. L., Freeman, M. R., & Mueller, P. S. (2011). Palliative care and end-of-life issues in patients treated with left ventricular assist devices as destination therapy. *Current Heart Failure Reports*, *8*(3), 212–218. doi:10.1007/s11897-011-0060-x PMID:21538039

Tilden, L. B., Williams, B. R., Tucker, R. O., MacLennan, P. A., & Ritchie, C. S. (2009). Surgeons' attitudes and practices in the utilization of palliative and supportive care services for patients with a sudden advanced illness. *Journal of Palliative Medicine*, *12*(11), 1037–1042. doi:10.1089/jpm.2009.0120 PMID:19663713

Truog, R. D., Campbell, M. L., Curtis, J. R., Haas, C. E., Luce, J. M., Rubenfeld, G. D., & Kaufman, D. C. et al. (2008). Recommendations for end-of-life care in the intensive care unit: A consensus statement by the American College [corrected] of Critical Care Medicine. *Critical Care Medicine*, *36*(3), 953–963. doi:10.1097/CCM.0B013E3181659096 PMID:18431285

Weissman, D. E., & Meier, D. E. (2011). Identifying patients in need of a palliative care assessment in the hospital setting: A consensus report from the Center to Advance Palliative Care. *Journal of Palliative Medicine*, *14*(1), 17–23. doi:10.1089/jpm.2010.0347 PMID:21133809

Wright, A. A., Keating, N. L., Balboni, T. A., Matulonis, U. A., Block, S. D., & Prigerson, H. G. (2010). Place of death: Correlations with quality of life of patients with cancer and predictors of bereaved caregivers' mental health. *Journal of Clinical Oncology*, *28*(29), 4457–4464. doi:10.1200/JCO.2009.26.3863 PMID:20837950

Wunsch, H., Christiansen, C. F., Johansen, M. B., Olsen, M., Ali, N., Angus, D. C., & Sorensen, H. T. (2014). Psychiatric diagnoses and psychoactive medication use among nonsurgical critically ill patients receiving mechanical ventilation. *Journal of the American Medical Association*, *311*(11), 1133–1142. doi:10.1001/jama.2014.2137 PMID:24643603

Wunsch, H., Guerra, C., Barnato, A. E., Angus, D. C., Li, G., & Linde-Zwirble, W. T. (2010). Three-year outcomes for Medicare beneficiaries who survive intensive care. *Journal of the American Medical Association*, *303*(9), 849–856. doi:10.1001/jama.2010.216 PMID:20197531

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Zier, L. S., Sottile, P. D., Hong, S. Y., Weissfield, L. A., & White, D. B. (2012). Surrogate decision makers' interpretation of prognostic information: A mixed-methods study. *Annals of Internal Medicine*, *156*(5), 360–366. doi:10.7326/0003-4819-156-5-201203060-00008 PMID:22393131

ADDITIONAL READING

Aslakson, R., Cheng, J., Vollenweider, D., Galusca, D., Smith, T. J., & Pronovost, P. J. (2014). Evidence-based palliative care in the intensive care unit: A systematic review of interventions. *Journal of Palliative Medicine*, *17*(2), 219–235. doi:10.1089/jpm.2013.0409 PMID:24517300

Baggs, J. G., Norton, S. A., Schmitt, M. H., Dombeck, M. T., Sellers, C. R., & Quinn, J. R. (2007). Intensive care unit cultures and end-of-life decision making. *Journal of Critical Care*, *22*(2), 159–168. doi:10.1016/j.jcrc.2006.09.008 PMID:17548028

Bradley, C. T., & Brasel, K. J. (2009). Developing guidelines that identify patients who would benefit from palliative care services in the surgical intensive care unit. *Critical Care Medicine*, *37*(3), 946–950. doi:10.1097/CCM.0b013e3181968f68 PMID:19237901

Campbell, M. L., & Guzman, J. A. (2003). Impact of a proactive approach to improve end-of-life care in a medical ICU. *Chest*, *123*(1), 266–271. doi:10.1378/chest.123.1.266 PMID:12527629

Campbell, M. L., & Guzman, J. A. (2004). A proactive approach to improve end-of-life care in a medical intensive care unit for patients with terminal dementia. *Critical Care Medicine*, *32*(9), 1839–1843. doi:10.1097/01.CCM.0000138560.56577.88 PMID:15343010

Carlet, J., Thijs, L. G., Antonelli, M., Cassell, J., Cox, P., Hill, N., & Thompson, B. T. et al. (2004). Challenges in end-of-life care in the ICU. Statement of the 5th International Consensus Conference in Critical Care: Brussels, Belgium, April 2003. *Intensive Care Medicine*, *30*(5), 770–784. doi:10.1007/s00134-004-2241-5 PMID:15098087

Curtis, J. R., Engelberg, R. A., Wenrich, M. D., Shannon, S. E., Treece, P. D., & Rubenfeld, G. D. (2005). Missed opportunities during family conferences about end-of-life care in the intensive care unit. *American Journal of Respiratory and Critical Care Medicine*, *171*(8), 844–849. doi:10.1164/rccm.200409-1267OC PMID:15640361

Curtis, J. R., Nielsen, E. L., Treece, P. D., Downey, L., Dotolo, D., Shannon, S. E., & Engelberg, R. A. et al. (2011). Effect of a quality-improvement intervention on end-of-life care in the intensive care unit: A randomized trial. *American Journal of Respiratory and Critical Care Medicine*, *183*(3), 348–355. doi:10.1164/rccm.201006-1004OC PMID:20833820

DeCato, T. W., Engelberg, R. A., Downey, L., Nielsen, E. L., Treece, P. D., Back, A. L., & Curtis, J. R. et al. (2013). Hospital variation and temporal trends in palliative and end-of-life care in the ICU. *Critical Care Medicine*, *41*(6), 1405–1411. doi:10.1097/CCM.0b013e318287f289 PMID:23518869

Goldsmith, B., Dietrich, J., Du, Q., & Morrison, R. S. (2008). Variability in access to hospital palliative care in the United States. *Journal of Palliative Medicine*, *11*(8), 1094–1102. doi:10.1089/jpm.2008.0053 PMID:18831653

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Hartog, C. S., Peschel, I., Schwarzkopf, D., Curtis, J. R., Westermann, I., Kabisch, B., & Reinhart, K. et al. (2014). Are written advance directives helpful to guide end-of-life therapy in the intensive care unit? A retrospective matched-cohort study. *Journal of Critical Care*, *29*(1), 128–133. doi:10.1016/j.jcrc.2013.08.024 PMID:24331948

Heyland, D. K., Dodek, P., Rocker, G., Groll, D., Gafni, A., Pichora, D.,... Canadian Researchers End-of-Life, N. (2006). What matters most in end-of-life care: perceptions of seriously ill patients and their family members. *CMAJ: Canadian Medical Association Journal, 174*(5), 627-633. doi: 10.1503/cmaj.050626

Johnson, S. K., Bautista, C. A., Hong, S. Y., Weissfeld, L., & White, D. B. (2011). An empirical study of surrogates' preferred level of control over value-laden life support decisions in intensive care units. *American Journal of Respiratory and Critical Care Medicine*, *183*(7), 915–921. doi:10.1164/rccm.201008-1214OC PMID:21037019

Kapo, J., & Siegel, M. D. (2013). Penguins and palliative care: Facilitating cultural change in the ICU. *Critical Care Medicine*, *41*(10), 2443–2444. doi:10.1097/CCM.0b013e31829136f3 PMID:24060776

Khandelwal, N., & Curtis, J. R. (2014). Clinical and economic implications of timing of limitations in life support. *Critical Care Medicine*, *42*(2), 455–456. doi:10.1097/CCM.0b013e3182a522de PMID:24434447

Lau, B. D., Aslakson, R. A., Wilson, R. F., Fawole, O. A., Apostol, C. C., Martinez, K. A., & Dy, S. E. et al. (2014). Methods for improving the quality of palliative care delivery: A systematic review. *The American Journal of Hospice & Palliative Care*, *31*(2), 202–210. doi:10.1177/1049909113482039 PMID:23532404

Lautrette, A., Darmon, M., Megarbane, B., Joly, L. M., Chevret, S., Adrie, C., & Azoulay, E. et al. (2007). A communication strategy and brochure for relatives of patients dying in the ICU. *The New England Journal of Medicine*, *356*(5), 469–478. doi:10.1056/NEJMoa063446 PMID:17267907

Lee Char, S. J., Evans, L. R., Malvar, G. L., & White, D. B. (2010). A randomized trial of two methods to disclose prognosis to surrogate decision makers in intensive care units. *American Journal of Respiratory and Critical Care Medicine*, *182*(7), 905–909. doi:10.1164/rccm.201002-0262OC PMID:20538959

Meissner, A., Genga, K. R., Studart, F. S., Settmacher, U., Hofmann, G., Reinhart, K., & Sakr, Y. (2010). Epidemiology of and factors associated with end-of-life decisions in a surgical intensive care unit. *Critical Care Medicine*, *38*(4), 1060–1068. doi:10.1097/CCM.0b013e3181cd1110 PMID:20068464

Mularski, R. A., Heine, C. E., Osborne, M. L., Ganzini, L., & Curtis, J. R. (2005). Quality of dying in the ICU: Ratings by family members. *Chest*, *128*(1), 280–287. doi:10.1378/chest.128.1.280 PMID:16002947

Nelson, J. E., Bassett, R., Boss, R. D., Brasel, K. J., Campbell, M. L., Cortez, T. B., & Weissman, D. E. et al. (2010). Models for structuring a clinical initiative to enhance palliative care in the intensive care unit: A report from the IPAL-ICU Project (Improving Palliative Care in the ICU). *Critical Care Medicine*, *38*(9), 1765–1772. doi:10.1097/CCM.0b013e3181e8ad23 PMID:20562699

Nelson, J. E., Cox, C. E., Hope, A. A., & Carson, S. S. (2010). Chronic critical illness. *American Journal of Respiratory and Critical Care Medicine*, *182*(4), 446–454. doi:10.1164/rccm.201002-0210CI PMID:20448093

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Nelson, J. E., Curtis, J. R., Mulkerin, C., Campbell, M., Lustbader, D. R., Mosenthal, A. C., & Weissman, D. E. et al.Improving Palliative Care in the. (2013). Choosing and Using Screening Criteria for Palliative Care Consultation in the ICU: A Report From the Improving Palliative Care in the ICU (IPAL-ICU) Advisory Board*. *Critical Care Medicine*, *41*(10), 2318–2327. doi:10.1097/CCM.0b013e31828cf12c PMID:23939349

Norton, S. A., Hogan, L. A., Holloway, R. G., Temkin-Greener, H., Buckley, M. J., & Quill, T. E. (2007). Proactive palliative care in the medical intensive care unit: Effects on length of stay for selected high-risk patients. *Critical Care Medicine*, *35*(6), 1530–1535. doi:10.1097/01.CCM.0000266533.06543.0C PMID:17452930

Penrod, J. D., Pronovost, P. J., Livote, E. E., Puntillo, K. A., Walker, A. S., Wallenstein, S., & Nelson, J. E. et al. (2012). Meeting standards of high-quality intensive care unit palliative care: Clinical performance and predictors. *Critical Care Medicine*, *40*(4), 1105–1112. doi:10.1097/CCM.0b013e3182374a50 PMID:22202706

Puntillo, K., Nelson, J. E., Weissman, D., Curtis, R., Weiss, S., Frontera, J., & Campbell, M. et al. (2014). Palliative care in the ICU: Relief of pain, dyspnea, and thirst--a report from the IPAL-ICU Advisory Board. *Intensive Care Medicine*, *40*(2), 235–248. doi:10.1007/s00134-013-3153-z PMID:24275901

The SUPPORT Principal Investigators. (1995). A controlled trial to improve care for seriously ill hospitalized patients. The study to understand prognoses and preferences for outcomes and risks of treatments (SUPPORT). *JAMA, 274*(20), 1591-1598.

Zalenski, R., Courage, C., Edelen, A., Waselewsky, D., Krayem, H., Latozas, J., & Kaufman, D. (2014). *Evaluation of screening criteria for palliative care consultation in the MICU: a multihospital analysis*. BMJ Support Palliat Care; doi:10.1136/bmjspcare-2013-000570

KEY TERMS AND DEFINITIONS

Consultative Model: A model of delivering palliative care using the services of a palliative care consultation service.

End-of-Life Care: Care, including pain and symptom management and family support, that is provided to patients at the end of their life, often when death is imminent.

Generalist Palliative Care: Basic palliative care, including pain and symptom management, prognosis and discussions regarding goals of care, that is to be performed by primary providers.

Integrative Model: A model of delivering palliative by incorporating principles of palliative care into routine ICU care.

Mixed Model: A model of delivering palliative care that combines the integrative and consultative models.

Palliative Care: A discipline of medicine that focuses on improving quality of life through the management of symptoms including pain, nausea, anxiety, psychological and spiritual distress for patients with serious illness.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Specialized Palliative Care: Advanced palliative care, including complex and refractory pain and symptom management, conflict resolution and complex decision-making that is performed by palliative care consultants.

Surgical Buy-In: A contract, often informally made, between a surgeon and a patient whereby both parties commit to undergo a surgical procedure and its attendant postoperative course.

Surgical Critical Illness: A trajectory of critical illness more commonly seen with surgical patients, where previously healthy patients suddenly become acutely and critically ill.

Triggers for Palliative Care Consultation: Screening criteria used to identify patients at high-risk of having palliative care needs for referral for palliative care consultation.

This work was previously published in Modern Concepts and Practices in Cardiothoracic Critical Care edited by Adam S. Evans, Gregory E. Kerr, Insung Chung, and Robin Varghese, pages 105-123, copyright year 2015 by Medical Information Science Reference (an imprint of IGI Global).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 2
An Introduction to the Basic Concepts in QSAR-Aided Drug Design

Maryam Hamzeh-Mivehroud
Tabriz University of Medical Sciences, Iran

Babak Sokouti
Tabriz University of Medical Sciences, Iran

Siavoush Dastmalchi
Tabriz University of Medical Sciences, Iran

ABSTRACT

The need for the development of new drugs to combat existing and newly identified conditions is unavoidable. One of the important tools used in the advanced drug development pipeline is computer-aided drug design. Traditionally, to find a drug many ligands were synthesized and evaluated for their effectiveness using suitable bioassays and if all other drug-likeness features were met, the candidate(s) would possibly reach the market. Although this approach is still in use in advanced format, computational methods are an indispensable component of modern drug development projects. One of the methods used from very early days of rationalizing the drug design approaches is Quantitative Structure-Activity Relationship (QSAR). This chapter overviews QSAR modeling steps by introducing molecular descriptors, mathematical model development for relating biological activities to molecular structures, and model validation. At the end, several successful cases where QSAR studies were used extensively are presented.

INTRODUCTION

Drug design and discovery is a multidimensional task necessitating collaboration across a broad range of disciplines and areas of research. Computational drug discovery is an essential component in drug design process that accelerates drug discovery and development by providing useful preliminary information at the early stages of the process. Quantitative structure-activity relationship (QSAR) is one of

DOI: 10.4018/978-1-5225-0549-5.ch002

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

the effective data mining strategies of the computational approaches which have opened up a number of informative perspectives in modern and profitable drug design process. The chapter will focus on rational drug design and the importance of relevant computational approaches in this area. Moreover, the chapter will introduce the QSAR studies as a part of computational methods in drug design and development processes. Different kinds of molecular descriptors, data analysis, descriptor selection, model building and evaluation will be discussed. Application of QSAR studies by providing some specific case studies will also be highlighted.

BACKGROUND

Drugs are vital, essential, and inevitable part of our life which is always being threatened by different diseases. Therefore, administration of safe and effective drugs is necessary. According to the side effects being reported in clinic, finding novel drugs with minimum toxicity is always a need. Drug discovery is a patient-oriented, complex, and time consuming process associated with spending huge amount of investment and involves many experts from different disciplines such as biology, biochemistry, pharmacology, mathematics, computing and molecular modeling. It has been estimated that of 9.000-10.000 new chemical entities identified or synthesized, one can reach the market within an average time of 16 years. Previously, drugs have been discovered either by identifying chemicals by trial-and-error or by serendipity. Today, attempts have been focused on rational drug design. Its fundamental principle lies in logical reasoning before synthesizing any therapeutic agent for the evaluation. In this context, computer-assisted drug design is a valuable and promising tool in rational drug design and discovery pipeline and plays an important role in pharmaceutical research by reducing the costly failures of drug candidates in clinical trials.

One of the strong appeals of QSAR studies deals with chemical safety and risk of toxicity. Toxicological profile and safety assessment of new chemical entities is of paramount importance in drug design and development process. Compounds with improved biological activity do not necessarily reach the market and most of them are discarded from clinical trials just for toxicity concerns. Therefore, toxicity is a major source of attrition for potential drug candidates. Drug toxicity and safety assessment of chemicals is time consuming and is accompanied with spending millions of dollars. In this context, structure-toxicity predictive models of pharmaceutically relevant molecules can provide an estimation of risk assessment and safety of drug candidates and it can be regarded as a promising and fundamental way to reduce the need for animal testing as well as ethical and monetary cost. As a consequence, potentially toxic compounds can be identified during early stages of drug discovery process. Carcinogenicity, mutagenicity, teratogenicity, hepatotoxicity, cardiotoxicity, and nephrotoxicity are the most important endpoints in toxicity evaluation (Sullivan, Manuppello, & Willett, 2014; Toropov, Toropova, Raska, Leszczynska, & Leszczynski, 2014).

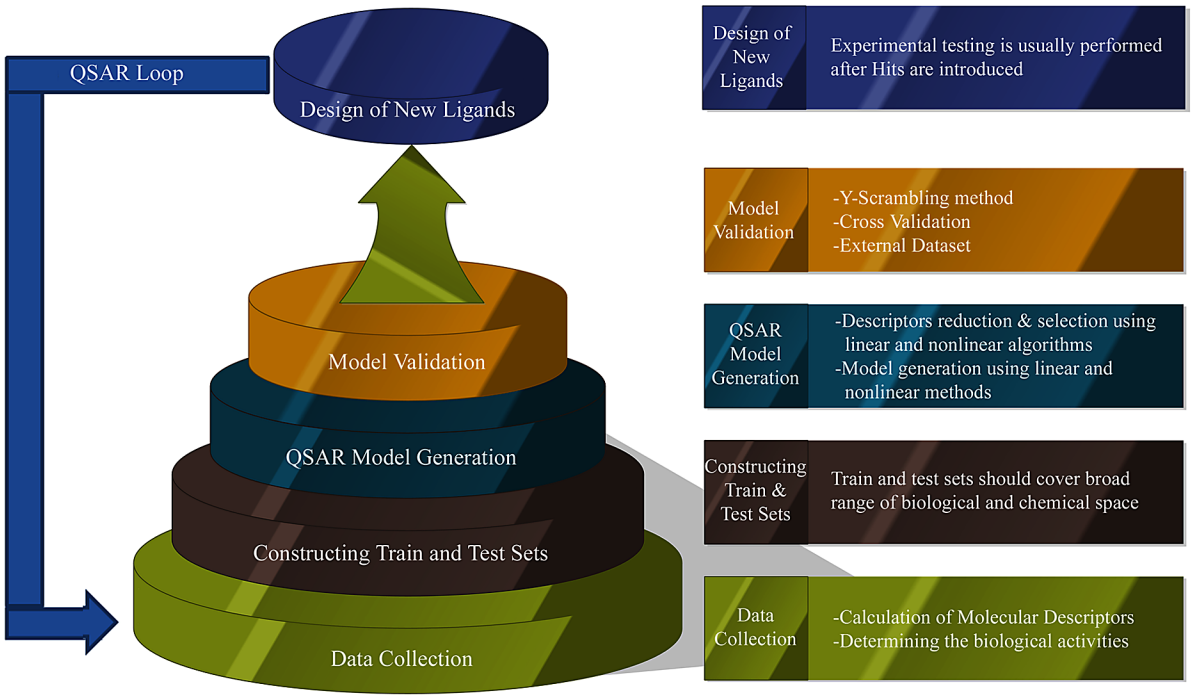
Computational approaches can be classified into structure-based drug design (SBDD), ligand-based drug design (LBDD) and sequence-based approaches (Ou-Yang et al., 2012). In the presence of experimentally determined structure of target molecule, molecular docking and *de novo* drug design can be used in the SBDD procedures. In situations in which three dimensional (3D) structure of target molecule is not available, LBDD methods are applied. Quantitative structure-activity relationship (QSAR), molecular field analysis, pharmacophore modeling, and 2D/3D similarity assessment are examples of these methods. In the absence of information regarding target and ligand molecules, sequence-based

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

approaches based on bioinformatics tools are utilized to ultimately find a solution to the drug development problem which may fit in one of the two previously mentioned categories.

QSAR methodology covers data analysis and statistical methods for developing models with capability of predicting the biological and toxicological activities of compounds based on their structures. It was first introduced by Corwin H. Hansch (Hansch, Maloney, Fujita, & Muir, 1962). The basic principle of the technique relies on building mathematical relationship between biological/toxicological activity data of a series of compounds and their structures defined by molecular descriptors. Molecular descriptors are numerical representations of molecular structures derived from steric, electronic, constitutional, topological, and geometrical properties of the molecules. Descriptors used in QSAR hold information about a molecule either as a whole or distinct parts of it. These descriptors can be experimentally measured or calculated using theoretical methods. Having descriptors in hand, those with biological relevance should be selected. Different descriptor selection methods, aimed to determine the best descriptor subset from a large pool of descriptors, have been proposed to reduce the number of descriptors present in the final mathematical QSAR model for explaining the biological activities of the studied molecules. These methods can be classified in three common categories including classical, artificial intelligence based, and miscellaneous methods that are necessary for generalization of predicted QSAR model through training steps. Multiple linear regression (MLR) analysis and partial least squares (PLS) are the examples of linear methods whereas artificial neural networks (ANNs) and the *k*-nearest neighbours (*k*-NNs) are few examples for non-linear methods. Generally, linear models predict the biological activity in a linear function and these models are also easy to interpret compared with non-linear models. Although, the classical methods can model the linear relationships between structural properties and biological activities,

Figure 1. Pyramid of different steps in QSAR procedure



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

however, they are not effective in developing models where the relationships are non-linear. Therefore, in these cases, the descriptor selection methods based on artificial intelligence can easily map existing non-linear relations between structural properties represented by descriptors and their corresponding biological activity faster and with higher reliability than classical methods because of their non-linear architecture. As QSAR methods are predictive modeling tools, the important criterion for evaluating their reliability and validity is to assess the capability of the models to predict not only the biological activity of the molecules included in training set (internal validation) but also for the molecules not involved in generating the model (external validation). Leave-one-out (LOO) and leave-group-out (LGO) are commonly utilized methods for internal cross validation whereas for external validation method, the predictivity of the QSAR model is judged based on a separate set of molecules called test set which are not included in model development in any stage. QSAR has served as a valuable predictive tool for identifying hit compounds, predicting the biological/toxicological activity of untested pharmaceuticals as well as optimizing the lead compounds for improving the biological activity (shown in Figure 1).

BIOLOGICAL ACTIVITY DATA

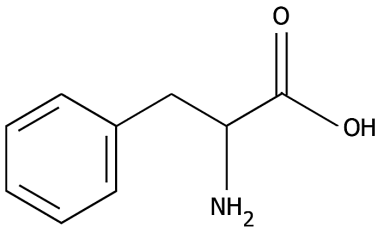
One of the essential constituents of QSAR analysis is biological activity of the compounds under investigation. It is defined as quantitative measurement of the changes brought about by any chemical entity in the biological process which can provide numerical assessment of potency (Jackson, Esnouf, Winzor, & Duewer, 2007). Experimental determination of precise and accurate biological activities is of great importance for generating a meaningful model as they affect the whole process of modeling down the way substantially. Biological activity obtained from a set of biological assays, should be performed with great care and evaluated by appropriate statistical treatment to the required satisfaction, e.g., doing the experiments in triplicates or doublet of triplicates due to response variations. A typical bioassay comprises two components including compound and target of interest like an enzyme, isolated receptors, animal tissue, cell based assays, and so on. Different kinds of biological and toxicological activities are used for QSAR studies including affinity data (substrate/receptor binding constant); rate constants such as association/dissociation and Michaealis Menten constants (K_d and K_m); inhibition constants like K_i and IC_{50} ; *in vitro* and *in vivo* pharmacokinetic and pharmacodynamic rate constants, lethal dose (LD_{50}) just to mention a few (Kubinyi, 1993). For QSAR analysis, equilibrium/rate constants are well suited as they are related to Gibbs free energy (ΔG ; $\Delta G = -2.303\ RT\ \log K_{eq}$; K_{eq} is the equilibrium constant). Therefore, the biological activity should be used in logarithmic scale to fulfill the linear free-energy relationship (LFER) criteria. The other reason is the normal distribution of the errors of the activity values when they are presented in the logarithmic scale. Sometimes, the logarithm of reciprocal biological activity is preferred to use as higher values obtained for potent analogs. The range of biological activity is important while developing the QSAR models. Full spectrum of activity is required for a data set and should range at least 3 to 4 orders of magnitude difference between the most and the least potent compounds. The larger the range, the more reliable QSAR model can be achieved (Scior et al., 2009). The compounds with poor activity should be incorporated in developing a predictive QSAR model. However, remotely isolated rare activities may be problematic by causing biases in the developed model. Since, the data for biological activity are crucial in generating accurate and valid QSAR model, therefore too much attention should be paid as they are not always as good as they seem at first sight.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

CHEMICAL STRUCTURE REPRESENTATION

There are several different ways of representing chemical structures including 2D and 3D representations. 2D structures specify elements, connectivity, and bond order whereas Cartesian coordinates for each atom is determined in 3D formats. Encoding approaches for 2D representations are line notations, connection tables and molecular graphs. Line notations are appeared as a sequence of atoms connected by compact linear string of alphanumeric symbols. The most commonly used line notations are described as follows. Simplified Molecular Input Line Entry System (SMILES; see Figure 2) is widely used format of notation implemented by Daylight Chemical Information Systems (Santa Fe, NM) (Weininger, 1988; Weininger, Weininger, & Weininger, 1989). SYBYL Line Notation (SLN) developed by Tripos[1] Inc. provides richer set of information for molecular queries, stereochemistry, and reactions (Figure 2). The other line notation introduced recently is InChI stands for IUPAC international chemical identifier. This format of representation provides a unique line notation consisting of six layers, each identifying a distinctive structural property about the molecule. The layers are main layer, a charge layer, stereochemical layer, an isotopic layer, a fixed-H, and a reconnected layer. The presence of main layer, which specifies atom and bond types, is required but the others are optional (Warr, 2011; Young, 2009). The other molecular line notations are Wiswesser Line-Formula Notation (WLN) and representation of structure diagram arranged linearly (ROSDAL). Figure 2 shows different line notation formats for phenylalanine as an example (Warr, 2011).

Figure 2. Different names and chemical representation as line notations for structure diagram of phenylalanine



Systematic name:	Phenylalanine
IUPAC name:	2-amino-3-phenylpropanoic acid
SMILES:	NC(Cc1ccccc1)C(O)=O
SLN:	C[1]H:CH:CH:CH:CH:C(:@1)CH2CH(NH2)C(=O)OH
WLN:	VQYZ1R
ROSDAL:	1O-2=3O, 2-4-5N, 4-6-7=-12-7
CAS RN:	150-30-1 (D/L)

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Connection tables are the other types of encoding style, in which the chemical structure is defined as a table of atom and bond types connecting the atoms with given arbitrary number in a series. Relative coordinates for each atom are usually included. An example of most commonly used connection table in Molfile format is shown in Figure 3.

Registry systems like Chemical Abstract Service Registry Number (CAS RN) are database of disclosed chemical substance information. Although it is a method of naming structures and provides no chemical information, however, it can be converted to machine readable structures by several programs such as ACD/Labs[2] (Toronto, Canada), CambridgeSoft[3] (Cambridge, MA), and InfoChem[4] (Munich, Germany) (Warr, 2011).

The other method for representing the chemical structures in 2D format is the molecular graph. Molecular graphs are derived from the graph theory; a theory not only used in chemistry but also in computer and related sciences. This theory is a branch of mathematics that defines the sets of objects by nodes and the corresponding relationships between them called edges. Atoms and bonds are considered as nodes and edges, respectively, in a molecular graph. For a molecular graph some properties like atom type (for nodes) and bond order (for edges) can be represented. Generally, the molecular graph shows the topology of the molecule and the way the atoms are connected (Figure 4).

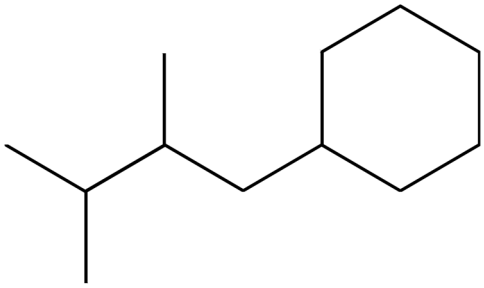
Figure 3. Molfile format for phenylalanine as an illustrative example of connection table style representation of molecules

```
Phenylalanine
  SLPMS1  03010512002D

 12 12  0  0  0  0  0  0  0  0999 V2000
    0.0000    2.0000    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -0.8660    1.5000    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -0.8660    0.5000    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    0.0000    0.0000    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    0.8660    0.5000    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    0.8660    1.5000    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    1.7321    2.0000    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    2.5981    1.5000    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    2.5981    0.5000    0.0000 N   0  0  0  0  0  0  0  0  0  0  0  0
    3.4641    2.0000    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    4.3301    1.5000    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
    3.4641    3.0000    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
  1  2  2  0  0  0  0
  2  3  1  0  0  0  0
  3  4  2  0  0  0  0
  5  4  1  0  0  0  0
  6  5  2  0  0  0  0
  1  6  1  0  0  0  0
  6  7  1  0  0  0  0
  7  8  1  0  0  0  0
  8  9  1  0  0  0  0
  8 10  1  0  0  0  0
 10 11  1  0  0  0  0
 10 12  2  0  0  0  0
```


EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 4. Molecular graph representation for phenylalanine



Although 2D representation of molecules provides chemical information, but in reality, molecules are 3D objects, and their atoms occupy distinct positions in a 3D space identified by the Cartesian co-ordinates. Some structural properties are obtained from 3D structures. The 3D structures are generated through empirical and numerical methods. X-ray crystallography, 2D NMR, IR, electron diffraction, and microwave spectroscopy are examples for experimental 3D structure determination. In numerical methods, quantum mechanics (QM)/molecular mechanics (MM) calculations are used for representing conformers of molecules. The 3D structure of a molecule is used for many different purposes, such as calculation of molecular descriptors, performing docking experiments, identification of pharmacophore, and building pharmacophore based 3D queries for similarity search in databases.

MOLECULAR DESCRIPTORS

Molecular descriptors (or sometimes called molecular parameters), the quantitative depictions of molecular structures, are crucial part of any QSAR study and encode certain structural features of molecules. The nature of descriptors varies according to the molecular properties. One may roughly classify descriptors as physico-chemical parameters (hydrophobic, electronic and steric properties), geometrical descriptors (molecular surface area calculation), constitutional descriptors (frequency of occurrence of substructure), electronic descriptors (molecular orbital calculation), and topological descriptors (connectivity index), however, as it is obvious they are to some extent interrelated. From dimensionality point of view, molecular descriptors are classified as one-dimensional (1D) representing bulk properties of compounds, two-dimensional (2D) topological and charge indices, three-dimensional (3D) descriptors indicating conformational aspects of a molecule. Parameters are determined for a fragment of a molecule (called fragment-based parameters/substituent constants) or calculated for a molecule as a whole (global descriptors) (Dutt & Madan, 2012; Estrada & Molina, 2001). In the following section these parameters are described briefly.

Fragment-Based Descriptors

Early fragment-based parameters were originated from physical organic chemistry studies where Hammett equation was developed based on rate and equilibrium constant of reaction of *meta*- and *para*- substituted benzoic acid (Hammett, 1937). These parameters refer to a part of molecule. In other words, they are resulted from the various functional groups on a fixed core structure of a homologous series. Therefore,

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

different substitutions on molecules have different impact on biological activities. For each substituent, a constant is defined as substitution constant based on the corresponding properties. Electronic, hydrophobic, and steric aspects of molecules are important properties that can be used for determination of different substituent constants in a series of compounds.

Electron density distribution in a molecule can be influenced by different substituents leading to change in biological activity of the compounds. Electronic constants indicate the polarity of the substituents. A good example of electronic constant is Hammett substituent constant (σ), as stated above. It measures the substituent's electron withdrawing/donating ability once substituted on benzoic acid defined by the following equation:

$$\sigma_X = \log(\frac{K_X}{K_H}) \tag{1}$$

In this equation, K_X and K_H indicate the acidic dissociation constants of the substituted and unsubstituted benzoic acid in aqueous solution. Electron withdrawing substituents will have positive values and vice versa.

Since, the lipophilicity of the molecules is of great importance in receptor interaction and pharmacokinetic behavior of drug candidates, therefore, the other substituent parameter derived using a similar concept to that of Hammett constant is hydrophobic constant (π). This parameter shows hydrophobicity of a substituent relative to hydrogen. It can be represented as follows:

$$\pi_X = \log \mathrm{P}_{RX} - \log \mathrm{P}_{RH} \tag{2}$$

where RX and RH denote the substituted and parent molecules, respectively; and P shows the lipophilicity. Positive value of π for a substituent shows that this substituent is more hydrophobic than hydrogen whereas the less hydrophobic substituent is characterized by negative value.

Sometimes, the size and shapes of substituents are dominant factors in biological activity of compounds. Quantification of steric substituent constants is accomplished by various steric parameters. Taft's steric parameter (E_s) was the first parameter describing the steric effect in QSAR studies (Taft, 1952). It is based on acid catalyzed ester hydrolysis of $RCOOR'$ *vs.* CH_3COOR' shown in the following equation:

$$E_s = \log(\frac{k_R}{k_{Me}})_A \tag{3}$$

where k_R and k_{Me} are the rate constants of hydrolysis for $RCOOR'$ and CH_3COOR'. The other steric substituent parameter was Charton's steric constant (υ) based on minimum van der Waals radius of a substituent. STERIMOL parameters were the other important steric features developed by Verloop (Verloop, 1987). These types of parameters are defined by spatial arrangement of a substituent in a molecule and obtained from distance-based measurement of the van der Waals and covalent radii along

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

with standard bond angles and lengths. For more comprehensive information see (Todeschini & Consonni, 2000; Verloop, 1987).

There are numerous examples in QSAR studies which are based on the fragment-based descriptors. In 2009, Du et al. developed fragment-based (FB) QSAR in which molecular scaffold of studied compounds was divided into several fragments with respect to substituents. Then, the biological activities were correlated with physicochemical properties of the molecular fragments using linear free energy equation by obtaining two sets of coefficients, one for weighting factors and the other for physicochemical properties of molecular fragments (Du et al., 2009).

Katritzky and colleagues proposed models for rate of skin permeability for a set of diverse compounds. In this study, substructural molecular fragments (SMF) were generated by splitting the molecular graphs to subgraphs for calculating their contribution for property of interest. Two different types of fragments were utilized including "sequences" and "augmented" atoms. The former one is related to sequences of atoms and bonds while the latter one represented a selected atom with its environment containing neighboring atoms and bonds with taking into the account the hybridization of the augmented atom (Katritzky, Dobchev, et al., 2006). The same approach was conducted for prediction of blood/tissue–air partition coefficient and blood–brain penetration for a set of organic compounds and drug molecules, respectively, using structural fragmental descriptors (Katritzky et al., 2005; Katritzky, Kuanar, et al., 2006).

Whole Molecule Descriptors

As mentioned earlier, another way for calculating the molecular descriptors is achieved by measuring or calculating different properties for a molecule and relating them to its whole structure, and not just to a fragment of it. Some of the important whole molecular descriptors are described briefly below.

Lipophilicity Parameters

Lipophilicity or hydrophobicity is one of the most attractive parameters in QSAR studies. This physicochemical property plays pivotal role in drug solubility, absorption (membrane permeation), distribution, and receptor interactions. It is commonly obtained from the logarithm of the partition coefficient of the compound between *n*-octanol and water represented by LogP. The most routine experimental methods for measuring the LogP are shake-flask method and liquid chromatography method. It can also be calculated by different methodologies implemented in software such as DRAGON[5], and ACD/Labs[2] suite of programs. It is an additive constitutive parameter as proposed by Hansch (Hansch, et al., 1962), who demonstrated this additive nature by summing up the π substituents for the fragments of a molecule to calculate its lipophilicity in the form of LogP value. The lipophilicity contribution of different groups and substituents in a molecule has also been represented by Rekker's equation (Rekker & Mannhold, 1992):

$$\log \mathrm{P} = \sum a_i f_i \tag{4}$$

In this equation, a_i and f_i indicate the number of occurrences of the molecular fragment and lipophilicity contribution, respectively.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Topological Descriptors

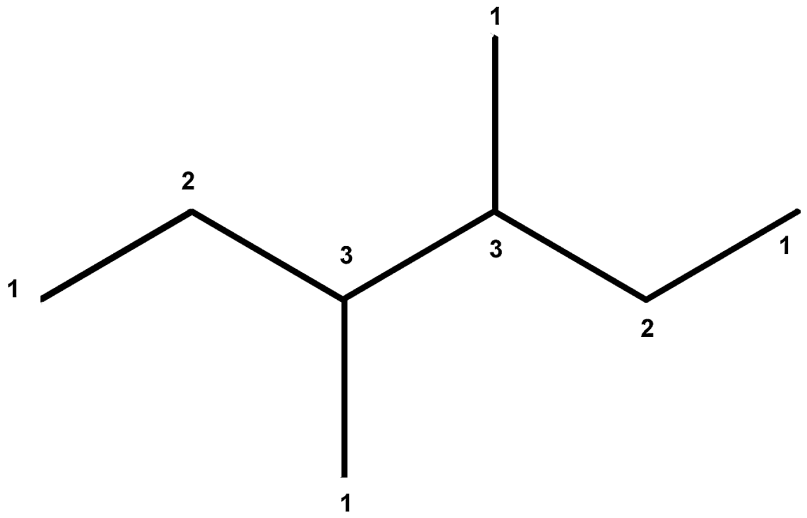
As the name implies, these descriptors/indices are derived from molecular topology i.e. molecular connectivity. They are conformationally independent and do not need the coordinates of atoms. Useful information about size, shape, degree of branching, presence of heteroatoms, and multiple bonds in a molecule are provided by these types of descriptors (Gozalbes, Doucet, & Derouin, 2002; Hall & Kier, 2001; Katritzky, Lobanov, & Karelson, 1995; Leach & Gillet, 2007; Randic, 2001). The Wiener index was the first topological descriptor developed which counts the number of bonds between pairs of atoms and sums the corresponding distance between all pairs (Wiener, 1947). The well known molecular connectivity indices are Randic index, and Kier & Hall index. The Randic index (RI) shows the branching index of a molecule. It is based on hydrogen suppressed graph representation of a molecule and defines a "degree" of adjacent nodes linked together by edges leading to calculation of bond connectivity value as follows:

$$RI = \sum_{allbounds} \frac{1}{(mn)^{1/2}} \tag{5}$$

where m and n denote the degree of atoms connected by bonds. Figure 5 shows an example of a connectivity index calculation for 3,4-dimethylhexane.

In the theory of Kier and Hall, the Randic index was modified and the number of hydrogens associated with atoms as well as valence electrons (sigma, pi and lone pair) were introduced (Kier & Hall, 1986). Subsequently, they designed the Kappa indices which are related to the molecular shape. These types of descriptors are not dependent on molecular geometry and obtained from possible "extreme shapes" for that number of atoms. For more detailed information see (Hall & Kier, 1991; Leach & Gillet, 2007).

Figure 5. Illustrative example of connectivity index calculation for 3,4-dimethylhexane



$$\mathbf{1}\chi = \frac{\mathbf{1}}{\sqrt{\mathbf{1}\times\mathbf{2}}} + \frac{\mathbf{1}}{\sqrt{\mathbf{2}\times\mathbf{3}}} + \frac{\mathbf{1}}{\sqrt{\mathbf{1}\times\mathbf{3}}} + \frac{\mathbf{1}}{\sqrt{\mathbf{3}\times\mathbf{3}}} + \frac{\mathbf{1}}{\sqrt{\mathbf{1}\times\mathbf{3}}} + \frac{\mathbf{1}}{\sqrt{\mathbf{2}\times\mathbf{3}}} + \frac{\mathbf{1}}{\sqrt{\mathbf{2}\times\mathbf{1}}} = \mathbf{3.717}$$

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Electrostatic Descriptors

A wide variety of descriptors can elucidate the electronic feature of the molecules which can be an important factor in determining the biological activity by affecting the strength of intermolecular interactions such as ion-ion, dipole-dipole, ion-dipole interactions as well as hydrogen bonding. Generally, these parameters reflect the polarity and ionization property of the molecule. The ability and tendency of a molecule for ionization can be exhibited via ionization constants (e.g., pK_a) which can provide some useful information about the extent of molecular ionization. This in turn influences the absorption, distribution, and excretion of the molecule. Another measure of polarity can be interpreted in terms of molar refractivity and molecular polarizability parameters which are highly interrelated (these parameters can be classified as steric parameters as well). Molar refractivity (MR) can be obtained by:

$$MR = \frac{n^2 - 1}{n^2 + 2}\frac{MW}{d} \tag{6}$$

In this formula, n denotes the refractivity index, MW and d are molecular weight and density, respectively. The positive value of MR in a QSAR model implies binding of a molecule to a polar surface whereas a negative value suggests an steric hindrance for binding. Refractivity index, itself accounts for polarizability of the molecule.

Quantum-chemical descriptors can most likely be classified in the category of electronic descriptors by providing valuable information about internal electronic feature of molecules. They are obtained by molecular orbital (MO) calculations. The most commonly used types of these descriptors are E_{HOMO} and E_{LUMO}, energies of highest occupied and lowest unoccupied MOs, respectively. The HOMO energy shows the ionization ability of the molecule and losing a pair of electrons to an electrophile while the LUMO energy is related to the magnitude of electron affinity of a molecule by accepting a pair of electrons from a nucleophile. Therefore, these descriptors determine the position and possibility of electrophilic and nucleophilic attraction of a molecule in drug-receptor interaction (Dastmalchi, Hamzeh-Mivehroud, Ghafourian, & Hamzeiy, 2008; Katritzky, et al., 1995). Dipole moment is the other parameter calculated from quantum mechanical calculations that estimates the polarity of the molecule and indicates the polar-type interactions between a molecule and its target (Karelson, Lobanov, & Katritzky, 1996).

Geometric Descriptors

Biological effect of a compound is the consequence of interaction between the molecule and its target. For a desired interaction, the degree of complementarity of compound and receptor is essential. Three-dimensional feature of a molecule structure can be represented by geometric descriptors. The atomic coordinates (x, y, z) of molecules are required for obtaining these descriptors and hence, they are influenced by molecular conformation. Therefore, size, shape, molecular conformation, and spatial arrangement of the atoms are important properties determined in geometric descriptors. Molecular volume, molecular surface area, shadow indices, and shape parameters are few examples of such descriptors (Katritzky & Gordeeva, 1993; Katritzky, et al., 1995).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Since the size of molecules is an important determining factor in drug-receptor interaction, molecular volume is regarded as a size related parameter which can be obtained from van der Waals volumes (Dudek, Arodz, & Galvez, 2006; Higo & Go, 1989) and used in QSAR studies. Molecular surface area is defined as amount of surface that is exposed to the solvent and is useful in prediction of drug transport, distribution, and absorption. Shadow parameters are calculated by projection of the molecular surface to its principal plane axes (i.e. XY , YZ , and XZ) using van der Waals radii for atoms (Todeschini & Consonni, 2000). The ratio of longest to shortest three dimensions of a rectangle having the shadows of molecule assuming van der Waals radii and standard bond lengths is defined as shape parameter (η) (Katritzky & Gordeeva, 1993; Todeschini & Consonni, 2000).

Constitutional Descriptors

These types of descriptors give mainly simple description of molecular composition without considering topologic, geometric, and electronic aspects of molecule. Molecular weight, number of atoms, atoms of different identity, bonds, functional groups, rings, and H-bond donors/acceptors are few examples. They are easily calculated in different programs like DRAGON[5] (Todeschini & Consonni, 2000). Although very simple, but when the feature being predicted shows significant correlation, these kinds of descriptors can appear in QSAR models.

DESCRIPTOR SELECTION METHODS

As discussed previously, the first stage for any QSAR study is to collect the data related to the molecular physicochemical properties. The collected data will include both the low-quality and high-quality data which can lead to low and high quality QSAR, respectively. However, the molecular descriptor selection from many existing types of molecular features is regarded as an important task in QSAR studies. For the purpose of removing low-quality or uninformative descriptors from the pool of molecular descriptors, many algorithms have been proposed to reduce the number of descriptors each representing a particular molecular feature. By omitting the redundant data, the next stage of QSAR study which is the building of QSAR model, can be carried out without interfering of noisy data and useless biases. Then, the pre- and post-processing techniques should be used to do the feature reduction and selection steps. The basic techniques for feature reduction in QSAR model building are partial least square (PLS) (Esposito Vinzi, 2010) and principle component analysis (PCA) (Ferraty & Romain, 2011; Jolliffe, 2002). The latter is a mathematical tool commonly used in cheminformatics to reduce descriptors from a high dimensional data space to a lower dimensional dataset known as orthogonal components by obtaining the transformed features which is generally a linear combination of the original descriptors. The resulted components are then sorted in descending order that finally, the component with the highest variance is selected as the first principal component to explain the most of the characteristics of the original data. In contrast to PCA, partial least squares, as a statistical method, constructs the components by maximizing the covariance between the response variable and a linear combination of predicting descriptors. However, the size of the reduced feature space achieved using PLS and the selected components are critical for classifiers' generalization problems (Li & Zeng, 2009). Recently, some of QSAR studies have been conducted based on PLS or PCA as a fundamental method along with other algorithms to improve the quality of the descriptor selection step. For example, ordered predictor selection (OPS)-PLS has been

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

used in a QSAR study for the selection of molecular descriptors important for antitrypanosomal activity of a series of thiosemicarbazones (Lozano et al., 2012).

The successive projection algorithm (SPA) is a variable selection tool consisting of two or more steps that can be used along with the linear discriminant analysis (LDA) (Goodarzi, Saeys, Araujo, Galvao, & Vander Heyden, 2013; Zhang, Rivard, & Rogge, 2008). By performing the SPA method, first, it identifies the most outstanding features and then the next distinct feature will be selected iteratively using the orthogonal projection to minimize the cost function by calculating the Mahalanobis distance values of molecules' features according to their true classification and misclassification. Gram-Schmidt orthogonalization (GSO) is also another pre-processing step for solving the collinearity problem to omit less informative descriptors by constructing a regression model between an starting descriptor and other descriptors to maximize the residuals as shown for 107 compounds with anti-HIV activity (Bagheri, Omidikia, & Kompany-Zareh, 2013).

Genetic algorithms (GA) known as evolutionary computation method inspired by gene duplication comprises five steps that is conducted iteratively to reduce the number of descriptors efficiently by parallel searching in a large scale data. These steps include the initialization (setting the population size), selection (a set of selection algorithms to select the features or descriptors that can be reproduced for the next generation), crossover (using numerous cross over algorithms such as 1-point, 2-point and uniform types for combination of selected descriptors to generate the offsprings based on probability P_c), mutation (this operator is generally applied after the cross over operator to mutate the offspring based on the probability P_m), and stopping criteria (the fitness function is needed to stop the iterative operation and converge the population globally) (Aguiar-Pulido et al., 2013). Finally, in the bit based GA, a "bit" translation of the existing features in QSAR study can be defined as active and inactive represented by "1" and "0" in a chromosome, respectively. GA has been implemented successfully in many QSAR studies (Goodarzi, Saeys, Deeb, Pieters, & Vander Heyden, 2013; Noorizadeh, Sajjadifar, & Farmany, 2013).

Simulated annealing (SA) (Oliveira Junior, 2012) is a probabilistic metaheuristic method for optimizing functions implemented in QSAR studies. This method is also based on evolutionary computations, which globally and iteratively searches in a large scale pool of descriptors to find the optimum descriptors with the least error function. Moreover, based on the probability values of Boltzmann distribution, the worst descriptor can also be selected which makes it hard for SA to be trapped in the local minima of error function. Ghosh and his colleagues have used SA effectively for descriptor selection in quinoxaline derivatives acting as anti-tuberculars (Ghosh & Bagchi, 2009).

Sequential feature forward selection (SFFS) (Sun, Todorovic, & Goodison, 2010) is a kind of classification tool which searches locally through the descriptors pool by using a sequential decision process. The SFFS method is used to assess the trained QSAR model based on the selected descriptors achieving a minimum error. At the first step, it is supposed that the best and the most informative descriptor is determined and selected. Then, the remaining descriptors will be used to evaluate the trained QSAR model based on the selected descriptors and the resulted minimum error. Finally, the descriptors with less error will be selected. As an example, a ligand-based virtual high-throughput screening with the PubChem[6] database has been carried out using the SFFS for assessing the problem-specific descriptor optimization protocols (Butkiewicz et al., 2013).

Sequential backward feature elimination (SBFE) (Sun, et al., 2010) is a type of sequential feature selection which acts in the reverse order to that of SFFS method. Despite the sequential forward feature selection which is based on adding features with minimum error, the sequential backward feature elimination is based on removing descriptors with maximum errors which is more time consuming. It selects

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

the worst and the most non-informative descriptors for being eliminated from the pool of descriptors. A novel method to overcome this disadvantage has also been presented by Guyon et al. which uses a recursive feature elimination based on support vector machines (SVMs) (Guyon, Weston, & Barnhill, 2002). Li et al. (Li, Cong, Yang, Xue, & Chen, 2013) have optimized their descriptors subset by recursive feature elimination method in an *in silico* study carried out on the spleen tyrosine kinase inhibitors.

Gravitational search algorithm (GSA) inspired from the Newton's gravity force, makes two massive objects move toward each other as any particle in the space absorbs the other particle (Bababadani & Mousavi, 2013). In this study, the binary based GSA is proposed which begins by generating a randomized binary agent as described in GA (i.e., an agent is a string of predefined variables which is a representative of absence or existence of descriptors and has a mass value calculated based on a fitness evaluation function). The GSA method has been employed for descriptor selection for anticancer potency modeling of a set of imidazo[4,5-b]pyridine derivatives in which the results showed that GSA outperforms the GA (Bababadani & Mousavi, 2013). However, this technique suffers from over-fitting problem and should be stopped to achieve the generalization property by using an algorithm such as regularization method.

The replacement method (RM) and its variations, with the property of not using the full search of descriptors, have been proved to achieve the optimal descriptors with a small number of linear regressions (Mercader, Duchowicz, Fernandez, & Castro, 2011; M. Sun et al., 2009). Moreover, the enhanced modified versions of RM resemble the simulated annealing techniques (Mercader, Duchowicz, Fernández, & Castro, 2008). The overall aim of RM is to choose an optimum set of descriptors with the minimum standard deviation which enables RM to find the global minimum. The RM method starts by randomly choosing an initial descriptor set and keeps the set with the minimum standard deviation. Then, all the descriptors with large standard deviation values will be chosen and substituted by remaining descriptors with small standard deviation. The process will be continued until the descriptors with minimum standard deviation are retained. However, in modified and enhanced editions of RM, this type of replacement is not performed (Sun, Zheng, Wei, Chen, Cai, et al., 2009). Some of QSAR studies related to the RM and its variants can be found in the literature (Mercader, Duchowicz, Fernandez, & Castro, 2010; Mercader, et al., 2011; Mercader, et al., 2008).

Cluster analysis is also another statistical method that separates the descriptors into individual classes using clustering algorithms such as k-means clustering method (Gonzalez, Teran, Saiz-Urra, & Teijeira, 2008). Most of QSAR studies are using the Fisher's discriminant ratio as their separation criterion (Lin, Li, & Tsai, 2004). After the criterion value is applied, the values of descriptors will be scaled in to a specified range of values (i.e., in most of the cases between 0 and 1). Then, the Fisher's discriminant ratio will be computed for all descriptors to choose the descriptors with the highest Fisher's discriminant ratio which will then be used to compute the cross-correlation coefficient between the descriptors with largest ratio and remaining descriptors. Cluster analysis method has been applied in some QSAR studies. For example, Prakash and Khan have used this method for identification of anticancer leads against human breast cancer (Prakash & Khan, 2013). QSAR studies on *Pseudomonas aeruginosa* deacetylase LpxC inhibitors, and HIV-1 integrase inhibitors are other examples (Kadam & Roy, 2006; Yuan & Parrill, 2005)

Ant colony optimization (ACO) algorithm, an stochastic optimization method, is generally based on the behavior of ants in selecting a path for finding the food which has the most pheromone with high probability rather than traveling randomly (Bonabeau, Dorigo, & Theraulaz, 2000). Some modified versions of ACO have also been proposed by authors to be applied to QSAR studies by simulating the ant colony systems. However, ACO is a time consuming process and like the linear methods suffers from trapping in local optima (Gonzalez, et al., 2008). The descriptor selection based on the ACO method is

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

generally carried out in a binary format (i.e., 1 and 0) in which for each variable, the existence and the absence of a descriptor is presented by 1 and 0, respectively. The amount of pheromone on each path will be updated according to an updating rule. Finally, ant selects a descriptor as its path based on the pheromone amount. The ACO process is iteratively done until a minimum error criterion is reached. However, the reported modified ACO techniques have advantages over the standard ACO on convergence speed and using fewer numbers of parameters (Gonzalez, et al., 2008). Modified ant colony optimization algorithm and its hybrid along with mutiple linear regression (MLR) have been used for variable selection in QSAR studies of cyclooxygenase inhibitors and tyrosine kinase inhibitor, respectively (Shamsipur, Zare-Shahabadi, Hemmateenejad, & Akhond, 2009; Shen, Jiang, Tao, Shen, & Yu, 2005; Shi, Shen, Kong, & Ye, 2007).

QSAR MODEL BUILDING METHODS

When in the preceding step, various sets of proper descriptors have been selected based on their corresponding molecular nature, different QSAR models can be generated using linear or non-linear methods (Aguiar-Pulido, et al., 2013; Doo Ho, Sung Kwang, Bum Tae, & Kyoung Tai, 2001). Although QSAR models are used for the prediction of the desired properties, however the predictions may sometimes fail to pass validation criteria even if a good training set has been chosen (Aguiar-Pulido, et al., 2013). QSAR modeling has become one of the most interesting research subjects because it may reduce the problems associated with the expensive synthetic methods, biological assays, and failure of drug developments due to poor ADMET profiles (Verma, Khedkar, & Coutinho, 2010). For this purpose, researchers have developed many QSAR software (Liew & Yap, 2012) ranging from structure drawing (e.g., ChemDraw[7], ACD/ChemSketch[8], OpenBabel[9]), 3D structure generation (e.g, CORINA[10], Concord[11], Frog[12], and smi23d[13]), descriptor calculation (e.g., ADRIANA Code[14], Dragon[5], Molconn-Z[15] and PaDEL-Descriptor[16]), and modeling (e.g., KNIME[17], RapidMiner[18], WEKA[19], Orange[20], TANAGRA[21], MATLAB[22] and R[23]) to multipurpose molecular modeling software packages including SYBYL[24], Discovery Studio[25], MOE (Molecular Operating Environment)[26], and CODESSA[27] developed by commercial vendors and academia. The QSAR modeling methodologies are divided into linear and non-linear methods on datasets with varying size, although some methods may be preferred for small datasets over the large datasets and *vice versa* (Aguiar-Pulido, et al., 2013).

Linear Methods

Using the linear methods implies that the function of QSAR model is comprised of linear combination of the molecular descriptors by using linear algorithms such as multiple linear regression (MLR), partial least square (PLS) or linear discriminant analysis (LDA) (Aguiar-Pulido, et al., 2013). By using MLR method, as one of the earliest methods used in building QSAR models, the linear function of biological activity is predicted based on a model with multiple independent descriptors by minimizing the squared deviations of the errors of the regression equation or maximizing the R^2 value by adjusting each of the available parameters (Afantitis et al., 2006). To reduce the correlations between the independent descriptors and/or eliminate the chance correlation, the number of independent descriptors are considered to be less than one-fifth the number of compounds in the training set, otherwise, multiple linear regression

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

method cannot be applied (Livingstone & Salt, 2005). An improved version of MLR known as Best MLR (BMLR) is used in various QSAR studies (Du, Wang, Hu, & Yao, 2008; Du, Zhang, Wang, Yao, & Hu, 2008) which overcomes most of the shortcomings of MLR method and makes it working well on training set less than five times the number of molecular descriptors (Liu & Long, 2009). Another advanced MLR method for linear QSAR model construction is the Heuristic Method (HM) (Katritzky, Lobanov, & Karelson, 2007) which can also be used to select descriptors for a non-linear model (Xia et al., 2009). Some advantages of HM are its high speed, no restrictions on the size of data, and good estimation of quality of correlations (Liu & Long, 2009). The overall descriptor selection in HM is comprised of ensuring about the availability of descriptor values in each structure and discarding the ones which are not available and/or significant for a specific structure, then calculating the pair correlation matrix for descriptors in order to keep only one of the highly correlated pairs of descriptors, and at the end gives estimates for the error of regression, cross-validated correlation coefficient, the F-test statistics, and the standard deviation used for evaluating the quality of correlation (Katritzky, et al., 2007).

The basic concepts of partial least squares (PLS) was proposed by Swedish statistician Herman Wold (Jores-Kong & Wold, 1982; Wold, 1966) to generate linear models from those with non-linear parameters by projecting the predicted variables and the observable variables to a new space. A PLS model, an extension of MLR, not only is used in descriptor selection step but also is used for QSAR model generation as well. It consists of three components including the latent variables, measurement, and weight components in which the weights are used to estimate the values of latent variables, so that the resulting values capture most of the variance of the independent variables that is useful for predicting the dependent variables. The overall goal is to use the descriptors to predict the responses (biological activities) in the population. This is achieved indirectly by extracting latent variables from sampled descriptors and responses, respectively. The extracted transformed descriptors are used to predict the responses. The PLS methodology has become popular in most QSAR studies related to CoMFA (comparative molecular field analysis) and CoMSIA (comparative molecular similarity index analysis), however its basic purpose was not for discriminative and classification problems (Barker & Rayens, 2003). A PLS regression can overcome the problems of standard regression when the matrix of predictors has more variables than observations, and when there is multicollinearity among descriptors (Olah, Bologa, & Oprea, 2004). Various combinations of PLS method with non-linear mathematical methods such as genetic function approximation (GA), factor analysis (FA), and orthogonal signal correction (OSC) have also been proposed to give better performance in QSAR analyses. The FA-PLS hybrid method employs FA for selecting the initial descriptors to find the relationships among variables in order to reduce the variables to be selected for the next step, i.e., PLS regression. A leave-one-out (LOO) method is mostly used for choosing the optimum number of components for PLS. Many examples related to the application of FA-PLS in QSAR studies have been presented, such as those used on HIV protease inhibitor mannitol derivatives, (phenylpiperazinyl-alkyl) oxindoles as selective 5-HT1A antagonists and cytochrome 3A4 inhibitory activity of structurally diverse compounds (Adhikari, Maiti, & Jha, 2010; Leonard & Roy, 2006; Roy & Pratim Roy, 2009).

GA-PLS is a combination of two calculation methods, namely, genetic algorithm (GA) (Maccari et al., 2006; Raichurkar, Shah, & Kulkarni, 2011; Williams et al., 2010) and PLS methods in which GA selects appropriate basis functions of a data model and PLS regression is used for setting the weights of the basis functions' relative contributions in the final model which in turn makes the model suitable in construction of larger QSAR equations.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Principal component analysis (PCA), a statistical procedure, was firstly invented by Karl Pearson (Pearson, 1901) and was later independently developed by Harold Hotelling (Hotelling, 1933) that uses orthogonal transformation to convert a set of observations of correlated variables into a set of values of linearly uncorrelated variables known as principal components for generating the predictive models. In this transformation, the first principal component has the largest possible variance, and each succeeding component has the highest variance such that it is orthogonal to the preceding components.

The independent property of principal components is guaranteed if the data set is normally distributed and the number of them is less than or equal to the number of original descriptors (Bagheri, Omidikia, & Kompany-Zareh, 2013). However, the structure of PCA is similar to that of LDA and FA on constructing the linear combinations of variables which best fit the data (Martinez & Kak, 2001).

Goodarzi et al. built their QSAR model based on PCA-ANFIS, a hybrid methodology consisting of the principal component analysis and the adaptive neuro-fuzzy inference systems for predicting the anti-HIV reverse transcriptase activities of tetrahydroimidazo [4,5,1-jk][1,4] benzodiazepine derivatives (Goodarzi & Freitas, 2010). In another study, PCA was employed to generate a robust model to predict the Src kinase inhibitory activity of 4-anilino-3-quinolinecarbonitriles (Sun, Zheng, Wei, Chen, & Ji, 2009).

Non-Linear Methods

The techniques discussed above were based on linear methods or a combination of them to develop the QSAR models. The linear methods may have a strong tendency to overfit the data once the large number of descriptors is used where the number of molecules is limited. In such cases, the non-linear nature of the problem requires the use of non-linear methods for the prediction of the activities based on the structural features of molecules. Modern QSAR models are based on the variations of the Hansch analyses that predicts a biological property as a statistical correlation with steric, electronic, and hydrophobic indices by using more powerful non-linear models, such as artificial neural networks (ANNs), support vector machines (SVMs), or other machine learning algorithms. The structural descriptors are numerical representations of some important molecular features, such as empirical indices (Hammett and Taft substituent constants), physical properties (octanol-water partition coefficient, dipole moment, aqueous solubility), counts of substructures or substituents, graph descriptors (Dury, Latour, Leherte, Barberis, & Vercauteren, 2001; Ivanciuc, 2003), topological (Fernandez-Blanco, Aguiar-Pulido, Munteanu, & Dorado, 2013; Fernandez-Lozano et al., 2013; Ivanciuc, 2013) or electrotopological (Sapre, Pancholi, Gupta, & Sapre, 2008; Wang, Liu, Shan, & Shi, 2010; Wang, Jiang, Pan, Cao, & Cui, 2009) and connectivity indices (Farahani, 2013; Mozrzymas, 2013), quantum chemical descriptors (atomic charges, HOMO and LUMO energies) (Angulo & Antolin, 2007; Hemmateenejad, Yousefinejad, & Mehdipour, 2011), and molecular interaction field descriptors (steric, electrostatic, and hydrophobic descriptors) (Baskin & Zhokhova, 2013; Gussregen et al., 2012). Some of these molecular parameters were described previously in Molecular descriptors section.

To find valuable drug candidates, many computational chemistry tools have been employed, among which artificial neural network (ANN) has a special place in modeling approaches because it provides solution to the complex problems with high accuracies.

The simulation of brain's neural function by variety of artificial neural networks (ANNs) was initially proposed by McCulloch and Pitts (Mc & Pitts, 1948; Pitts & Mc, 1947), and Rosenblatt (Rosenblatt, 1962). The basic structure of an ANN consists of input, hidden, and output layers, commonly structured in a multilayer feedforward (MLF) network usually trained by the backpropagation algorithm to model

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

the non-linear relationships between input variables and the output properties by adjusting their weights and biases. The system uses the mean squared error (MSE) through the training process. A special type of feedforward ANN is the radial basis function (RBF) neural network with only three layers, including input layer, a hidden layer and a linear output layer (Moody & Darken, 1989) which can predict wide range of properties, such as protein folding (Abbasi, Ghatee, & Shiri, 2013), protein interaction (Chen, Xu, Yang, Zhao, & He, 2012), and beta barrel membrane protein detection (Chen, Ou, Lee, & Gromiha, 2011; Ou, Gromiha, Chen, & Suwa, 2008).

Graph machine neural networks, which work based on the graph representation of the molecular structures, have been implemented in variety of computational chemistry tools with the capability of performing QSAR analyses (Baskin, Ait, Halberstam, Palyulin, & Zefirov, 2002; Chupakhin, Marcou, Baskin, Varnek, & Rognan, 2013) such as ChemNet (Kireeva, 1995), and MolNet (Ivanciuc, 1999).

Genetic algorithm (GA) is another type of biologically inspired algorithm initially proposed and developed by Holland (Holland, 1975) and Goldberg (Goldberg, 1989). GA is based on the Darwinian evolution of a population of individuals (i.e., chromosomes) to solve the high dimensional non-linear problems in which each chromosome is a solution to the problem and the chromosomes can be encoded as binary, continuous or a hybrid of both and their qualities are assessed by their fitness values. GA and its variants have been employed in many cheminformatics and computational chemistry applications. The followings are some demonstrated examples of its use; 3D-QSAR studies on novel phosphodiesterase-4 inhibitors, pharmacophore identification for benzotriazine derivatives, identification of novel inhibitors for cancer targets, acetylcholinesterase, HIV-1 protease, and ion-channels, as well as developing anti-protozoan compounds (Fernandez, Caballero, Fernandez, & Sarai, 2011; Pourbasheer, Riahi, Ganjali, & Norouzi, 2010; Sahin, Saripinar, Yanmaz, & Gecen, 2011). Evolutionary algorithms such as GA are also successfully applied to descriptor selection and global optimization of QSAR models (Perez-Castillo et al., 2012; Sahin, et al., 2011).

Ant colony optimization (ACO), an agent/colony based algorithm developed by Dorigo et al. (Dorigo, Maniezzo, & Colorni, 1996), simulates the foraging behavior of ants. The main aim of the ACO algorithm (i.e., the ant population) is to search and find the shortest path to the food resource by means of a chemical substance called pheromone accumulated on the paths that are more regularly explored by ants. The larger the amount of pheromone, the larger the probability for a path to be the shortest one to reach the food resource. The amount of pheromone in each path will be updated according to the amount of pheromone deposited by an ant. Like the GA, ACO method and its modified versions have also many applications in optimization problems such as chemistry and drug design where the target solution space is huge. There are several hybrid implementations including ACO-MLR (Goodarzi, Jensen, & Vander Heyden, 2012), modified ACO-neuro fuzzy interference for serotonin (5-HT7) receptor inhibitors (Jalali-Heravi & Asadollahi-Baboli, 2009), and ACO-random forest (Patil, Raj, Shingade, Kulkarni, & Jayaraman, 2009) for generating the QSAR models in which the main object is to cluster the new compound in large datasets of chemical compounds based on their similarities measured by molecular descriptors. The classical ACO can only be used for optimization problems related to discrete variables, whereas many QSAR studies need to optimize the continuous variables that has been addressed as an extension to ACO proposed by He and co-workers (He, Chen, & Zhao, 2006). Izrailev and Agrafiotis proposed two ACO approaches to find the best regression trees where each ant was an indicator for a regression tree, and to select required features by using ANN (i.e., ANTSELECT) on three QSAR datasets (Izrailev & Agrafiotis, 2001; Izrailev & Agrafiotis, 2002).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Particle swarm optimization (PSO), known as swarm intelligence (SI) proposed by Kennedy and Eberhart (Kennedy & Eberhart, 1995), is used for solving the non-linear optimization problems by simulating the social behaviors such as swarming, herding or flocking of various spices such as bees, birds, or fish started from a random position with random velocity in a search space that can be applied to both binary or continuous variables which is capable of fast convergence by using a small population and small number of iterations. The new positions in PSO algorithm are determined by either the particle's experience obtained from its previous movement or the best particle's memory in the swarm. That is why, in several cheminformatics studies, PSO and its modifications have been known as efficient replacements of GA and ACO for global optimization. Moreover, PSO has been successfully applied in many non-linear problems such as QSAR (Cheng, Zhang, & Zhou, 2011), protein motif discovery (Hardin & Rouchka, 2005), enzyme-inhibitor docking (Goodarzi, Saeys, Deeb, et al., 2013; Lu, Shen, Jiang, Shen, & Yu, 2004), protein-ligand geometry (i.e., Tribe-PSO in AutoDock) (Goodsell, 2009; Goodsell, Morris, & Olson, 1996; Morris, Goodsell, Huey, & Olson, 1996; Tiwari, Saxena, Pant, & Srivastava, 2012), and finding the global minimum geometry of chemical compounds. Based on the literature, the performance of PSO and its variants is promising, and they are popular function optimization methods in many machine learning algorithms.

STATISTICAL TREATMENTS AND MODEL VALIDATION

Data analysis and processing is usually accompanied by evaluation procedure. Evaluation of the developed models constitutes fundamental and inevitable step in QSAR studies and are paid particular attention. Any QSAR model should be evaluated prior to its use for data interpretation as well as prediction of biological activities. There are different tools to assess the quality, validity, and reliability of the generated QSAR models. The foremost judgment on a QSAR model is based on the statistical parameters of corresponding model. Squared correlation coefficient (R^2, r^2) is one the most commonly reported parameters in QSAR models that assesses goodness-of-fit and describes the degree of the variation in the variables explained by the regression equation ranging from zero to unity. The ideal and perfect model has an R^2 value of 1 explaining all the response (i.e., biological activity) whereas zero value of R^2 demonstrates no explanatory ability of the model. The other parameter is residual standard deviation (denoted by *s*) indicative of how well a model can predict the observed activities. Therefore, the minimum requirements for a good QSAR model is to have high R^2 (close to unity) and low *s* values.

The fisher statistics (F) is another QSAR statistical parameter for assessing the significance of the models by carrying out analysis of variance (ANOVA). It is described as the ratio of the explained mean square to the residual mean square. This term shows how well a regression equation fits to the training data set in a specified confidence interval. A significant QSAR model should have a high F value. It should be kept in mind that the best amount of statistical parameters mentioned above are not sufficient criteria for model reliability and validity and can not be considered as the indicators of predictive capability of the QSAR model. They just indicate how a model is capable to reproduce the biological activity (in this case) for the training data set.

As stated earlier there are several strategies for the validation of QSAR models such as,

1. Cross validation

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

2. External validation, and
3. Y-scrambling methods.

There are two types of cross validation (CV) methods namely, leave-one-out (LOO) and Leave-many/group/several-out (LMO) which evaluate the predictivity. They are known as "internal cross validation". In these methods one or different portion of compounds are removed in such a way that each data point is taken away once from the entire data set followed by iteratively fitting the model to the remaining data points and prediction of the biological activity of the left out molecules. Comparing the observed *vs* predicted values of the biological activities leads to calculation of the "cross validating R^2", which is routinely called Q^2 or q^2. q^2 is calculated according to the following equation (Golbraikh & Tropsha, 2002; Tropsha, Gramatica, & Gombar, 2003):

$$q^2 = 1 - \frac{\sum_{i=1}^{N}\left(y_i - \tilde{y}_i\right)^2}{\sum_{i=1}^{N}\left(y_i - \overline{y}\right)^2} = 1 - \frac{PRESS}{SSD} \tag{7}$$

Here SSD is the sum of squared deviations for each actual activity value y_i from the average activity $\overline{y}$, over the entire data set. $PRESS$, the predictive residual sum of squares, is the sum of squared differences between the actual activity y_i and the predicted activity $\tilde{y}_i$. N is the number of data points.

LOO is usually performed for small data set (e.g., 25 compounds), while for big data set, the LMO is preferred. Generally, LMO is considered as a stronger validation approach compared to LOO method due to the fact that in each cycle of LMO method constant perturbation is performed by eliminating certain number of data points while in LOO method by increasing the number of data points just $\frac{1}{N^{th}}$ of entire data set is removed. Consequently, smaller perturbation in data structure leads to obtaining q^2 values too close to R^2 (Gramatica, 2007; Wold & Eriksson, 1995). Although, q^2 determines goodness-of-prediction, however, it should not be regarded as merely criterion of the robustness and predictive ability of the model. Some QSAR practitioners consider the values of $q^2 > 0.5$ are good enough for validity of a model, but in reality, greater values for q^2 are essential conditions for validity of a model but not adequate. Golbraikh and Tropsha demonstrated that in addition to internal validation methods, further validation methods should be used for establishing a reliable QSAR model (Golbraikh & Tropsha, 2002). More discussion on this issue can be found elsewhere (Dearden, Cronin, & Kaiser, 2009; Gramatica, 2007; Tropsha, et al., 2003). Therefore, internal cross validation gives only an approximation of the accuracy of a QSAR model. To be more rigorous, the "external validation" approach must be employed for establishing a truly predictive QSAR model. To this end, a validation data set (i.e., data points not used in the model development procedure) is chosen and the model is used to predict their biological activities. The predictions are compared with actual activities and the agreement among them is evaluated. The higher the correlation between the predicted and the observed activities for the validation data set, the more robust is the model. Therefore, the entire available data set is split into training and test data sets. The important requirement is that both of data sets should span in such a manner that all ranges in terms of biological activity and structure are adequately covered and the number of test

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

data set should not be less than five (Golbraikh & Tropsha, 2002; Gramatica, 2007; Tropsha, et al., 2003). Generalizability of the developed QSAR model for new chemical entities is attained where the external validation step is carried out at the model development stage.

The last but not least validation strategy for evaluating robustness of the QSAR model is using "Y-scrambling" method (also called Y-randomization test). This technique is based on randomization of the responses (i.e. biological activities) in intact independent variables (i.e. molecular descriptors) followed by full data analysis using the scrambled activities. This process is performed repetitively for at least ten trials. The resultant statistics (i.e., R^2 and q^2) are expected to be very low compared to that of the original data. Otherwise, the model is not reliable and applicable (Golbraikh & Tropsha, 2002; Tropsha, et al., 2003; Wold & Eriksson, 1995).

For a predictive QSAR model, the following statistical criteria are also recommended (Golbraikh & Tropsha, 2002; Tropsha, 2010; Tropsha, et al., 2003):

1. $\mathrm{q}^2 > 0.5$
2. $R^2 > 0.6$
3. $\dfrac{(R^2 - R_0^2)}{R^2} < 0.1$
4. $0.85 \leq k \leq 1.15$

Here, q^2 is cross validated correlation coefficient, R^2 is squared correlation coefficient between predicted and observed activities, the correlation coefficients between predicted and observed activities R_0^2 when the regression line passes the origin, and *k* shows the corresponding slope. All of the criteria are calculated for test set compounds except the first one (i.e., q^2).

For developing a highly predictive model, serious concern should be paid during the evaluation step and a trustworthy QSAR model should encompass carefully the above outlined principles.

3D-QSAR STUDIES

The experimentally measured biological activity of the compounds arises from the interaction(s) between a molecule and its target. Therefore, determining the stereoelectronic properties accounting for ligand-target recognition process and the subsequent interaction(s) is the main goal of a three dimensional quantitative-structure activity relationship (3D-QSAR) study. This technique is introduced as an integration of classical QSAR with molecular modeling approaches. From computational point of view 3D-QSAR is more complex than classical QSAR and provides depiction of possible important interactions between a series of compounds and target of interest in a quantitative fashion. In other words, to fulfill these interactions, compounds should have a minimum structural requirement which is usually referred to as "pharmacophore". A pharmacophore abstracts the molecules into the functional groups capable of interacting with target structure. The structural descriptors used for 3D-QSAR studies are three dimensions in nature originated from measured interaction energies between a probe atom/group around a series of molecules.

The attractive advantage of 3D-QSAR relies on ability of predicting the interaction(s) even having no information about the active site of target that can be a basis for ligand-based drug design (LBDD).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

There are different steps for a 3D-QSAR approach. The first step is collecting the biological data and selecting the training set. The biological activities should have a wide range of activity as possible. The next step is conformational analysis of the molecules. In this step, the bioactive conformers are determined. Bioactive conformers refer to the conformation of the molecules that bind to the receptor. This kind of information is achievable both by experimental and theoretical techniques. The useful experimental data can be provided by X-ray crystallography and NMR spectroscopy (Verma, et al., 2010). In the case of theoretical approaches where the 3D structure of receptor is unknown, the comparative modeling can be applied to develop most probable model structure of the target for evaluation of ligand-receptor interactions. In order to determine the bioactive conformations of the ligands some criteria are considered for example, the energies of all conformations should be near the lowest energy and all training set molecules should be in similar conformations where the pharmacophoric components are aligned (Young, 2009). Bioactive conformers are aligned/superimposed over each other. Then, the molecular interaction fields (MIF), which will be explained in more details below, are calculated to be used as molecular descriptors. Different multivariate statistical techniques such as partial least square (PLS) and principal component analysis (PCA) are utilized for model generation by extracting the essential information from large pool of descriptors for the purpose of correlating with biological activity. Then, favorable and unfavorable regions for biological activities are visualized. The developed model is evaluated using cross validation methods and can be interpreted for subsequent important applications such as activity prediction for new compounds.

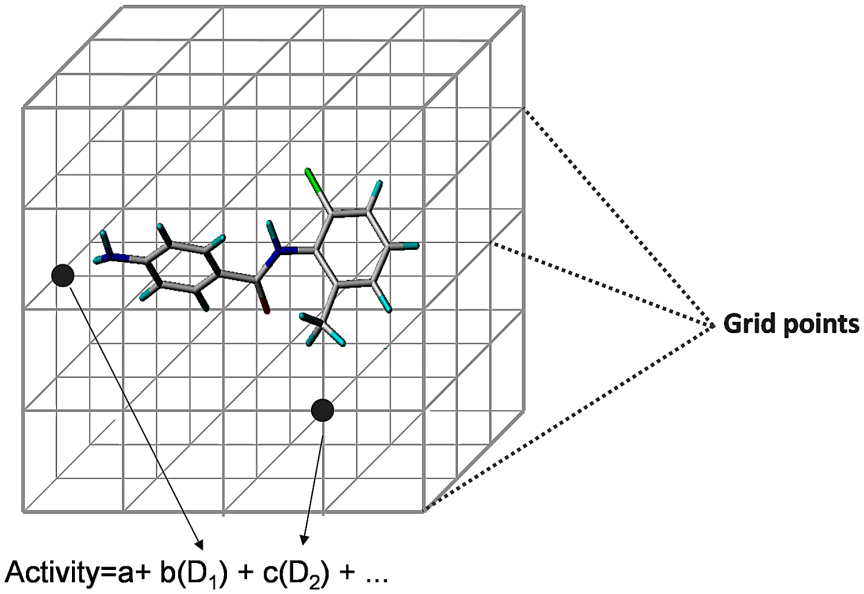
Generally, for classification of 3D-QSAR methods one may follow two approaches: alignment-dependent methods and alignment-independent methods. In the former, the calculation of descriptors is strongly influenced by the alignment of the molecules guided by experimental data or computational methods whereas in the latter case, the superposition of molecules is not required.

The most well-known and prototype of alignment-dependent methods is comparative molecular field analysis (CoMFA) which is introduced in 1988 (Cramer, Patterson, & Bunce, 1988). It is based on the calculation of electrostatic and steric interaction energies (i.e. MIFs) between a probe and a series of molecules (Dudek, et al., 2006). MIF explains the interaction energy variation between a selected probe and a set of molecules that are previously overlaid over each other and positioned at the center of a 3D grid box (Cruciani, 2006) (shown in Figure 6). A probe is a small molecule like water or a specific fragment such as methyl placed in a lattice of grid points. Each grid point (intersection of the lattice) defines a point in space and becomes a descriptor variable in 3D-QSAR. In the case of CoMFA, they are specified as electrostatic or steric descriptors. Several force fields at different level of approximation are utilized for determination of interaction energies. A force field represents a mathematical equation containing standard bond lengths, bond angels, dihedral angles, interatomic distances and the other parameters. The standard (6-12) Lennard-Jones function and Coulomb's law are employed in the force fields of CoMFA for computation of interaction energies (Kerdawy, Gussregen, Matter, Hennemann, & Clark, 2013; Verma, et al., 2010).

Comparative molecular similarity indices analysis (CoMSIA) is the other 3D-QSAR alignment-dependent method based on similarity indices of the molecules. It is too similar to CoMFA. Steric, electrostatic, hydrophobic and hydrogen bonding properties are considered for calculating descriptors using Gaussian-type potential function (Klebe, Abraham, & Mietzner, 1994). One of the advantages of CoMSIA, is that it provides the accurate information in grid points positioned within the molecule (Dudek, et al., 2006; Munteanu et al., 2010; Verma, et al., 2010). Both the CoMFA and the CoMSIA are implemented in SYBYL[7] program, Tripos[1], Inc.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 6. Molecular interaction field calculation based on interaction energy between the probe molecule (i.e., circles) and the molecule in a 3D grid box; each grid point represents a descriptor variable in 3D QSAR analysis.



The GRID[28] approach, introduced by Goodford in 1985 (Goodford, 1985), is another useful method in 3D-QSAR studies. The methodology proceeds by calculation of total interaction energies between different chemical probes and data set. GRID determines energetically favorable binding sites on known molecular structures as well as surface properties. In addition to steric and electrostatic potentials, hydrogen bonding and hydrophobic potentials are also included. The major advantage of this method is that GRID potentials have been developed based on experimentally determined ligand-protein complexes. Moreover, the availability of wide variety of probes provides the modeling of different types of interactions (Cruciani, 2006; Verma, et al., 2010). GRID is supplied by Molecular Discovery[29], Ltd.

The above mentioned 3D-QSAR methods rely on alignment step, however, there is other approach where the 3D-QSAR models are developed free from the alignment of studied compounds. Alignment-independent methods eliminate the superpositioning process which can be an advantage compared to the alignment-dependent methods. One of the examples of alignment-independent 3D-QSAR methods is comparative molecular moment analysis (CoMMA). In this method second-order moments of molecular mass and charge distribution are considered for generating the molecular similarity descriptors (Silverman & Platt, 1996). These kinds of descriptors are insensitive to the molecular alignment and are sensitive to the molecular conformation (Dudek, et al., 2006; Verma, et al., 2010).

GRid-INdependent Descriptors (GRIND) comprise a class of alignment-independent descriptors that were developed by Pastor *et al.* in 2000 (Pastor, Cruciani, McLay, Pickett, & Clementi, 2000). In GRIND method, the descriptors are obtained from an MIF collection calculated by probing of grid with chemical probes as before and then, the most favorable energies of interactions (hot spots) are selected. This is followed by encoding the relative position of specified regions into a group of values characterizing hot spots positioned at distinct distance ranges (Cruciani, 2006; Dudek, et al., 2006; Duran, Zamora, & Pastor, 2009). GRIND method is implemented in Pentacle[30] program, Molecular Discovery Ltd[29].

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

These various methods outlined above are just few examples of two approaches (namely alignment dependent and alignment independent methods). More information about the other different approaches can be found in the literature (Akamatsu, 2002; Cruciani, 2006; Datar, Khedkar, Malde, & Coutinho, 2006; Dudek, et al., 2006; Gohlke & Klebe, 2002; Lushington, Guo, & Wang, 2007; Polanski, Gieleciak, & Bak, 2002; Polanski & Walczak, 2000; Verma, et al., 2010; Walters & Hinds, 1994).

The application of multivariate chemometric techniques is inevitable because of the huge number of generated descriptors, as the traditional linear regression analysis can not be used in these cases. Partial least squares (PLS) and principal component analysis (PCA) are extensively used for data analysis, by correlation of 3D-descriptors (dependent variables) with related biological activities (independent variables) assuming a linear relationship among them. Whenever, the relationship is non-linear, chemometric methods such as artificial neural network is preferred.

An advanced PLS method called GOLPE (Generating optimal linear PLS)[31] was introduced in 1993, with the aim of developing highly reliable models with improved predictivity. In this method, fractional factorial design (FFD) is primarily used for combination of variables and in this process only variables which increase the predictive power of the models are selected for final PLS and classified according to their predictivity contribution (Baroni et al., 1993).

For evaluation procedure cross validation methods are applied as explained in Statistical treatment and model validation section. Leave-one-out and leave-group-out are used for internally validating the model although for further validation, external validation is strongly recommended.

Multi-Dimensional QSAR

The most pronounced recent advance in QSAR methodology was the expansion of its dimensionality. These multi-dimensional methods are steadily establishing their own place in QSAR studies by incorporating extra dimensions, such as physical characteristics, to overcome the obstacles of the existing QSAR methods. Multi-dimensional QSAR approaches are extended forms of 3D-QSAR methods in which additional multivariate molecular descriptors derived from 3D molecular representations, are included. Table 1 shows different multi-dimensional QSAR methodologies and their properties. Hopefinger and his coworkers were the pioneers of 4D-QSAR development (Hopfinger et al., 1997). In 4D approach, the "fourth dimension" also called "ensemble sampling" is determined by including conformational flexibility and the freedom of alignment by ensemble sampling the spatial features such as conformations, orientations, and protonation states for each molecule (Andrade, Pasqualoto, Ferreira, & Hopfinger, 2010; de Melo & Ferreira, 2012; Martins, Barbosa, Pasqualoto, & Ferreira, 2009; Polanski, 2009; Vedani, Dobler, Dollinger, et al., 2005). In this method, the bias associated with the choice of bioactive conformation is reduced (Vedani, Dobler, Dollinger, et al., 2005). 4D-QSAR analysis can be used as receptor independent (RI) and receptor dependent (RD) methods. In RI-4D-QSAR method the maximum structural information is obtained through pharmacophoric feature of substituents on the ligands' scaffold. When the geometry of receptor is known, RD-4D-QSAR analysis is utilized for mapping the ligand-receptor interaction (Damale, Harke, Kalam Khan, Shinde, & Sangshetti, 2014; Polanski, 2009). An example of 4D-QSAR technique, which has been developed based on CoMFA method is "LQTA-QSAR" (LQTA, *Laborato´rio de Quimiometria Teo´rica e Aplicada*). In this method, conformational ensemble profile is generated for each compound followed by obtaining the 3D descriptors through calculation of intermolecular interaction energy at each grid point considering probes and all aligned conformations obtained from molecular dynamics (MD) simulation (Damale et al., 2014; de Melo &

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Ferreira, 2012; Martins et al., 2009). Further progress in adding dimensionality to the QSAR techniques was brought about by considering the changes in the conformation of the receptor upon ligand binding (referred to as induced fit effect). Induced fit effect is considered as adaptation of the binding pocket to the individual ligand topology (Damale et al., 2014; Vedani, Descloux, Spreafico, & Ernst, 2007; Vedani & Dobler, 2002; Vedani, Dobler, Dollinger, et al., 2005). This topology alteration in the binding site through induced fit effect may influence the properties of the binding pocket such as hydrophilic, hydrophobic, and dielectric characteristics (Vedani et al., 2007). Addition of another dimension leads to development of 6D-QSAR which takes into account the solvation function as well as those already covered up to the fifth dimension. In 6D-QSAR, different solvation models are incorporated by mapping the surface area with different solvent properties (Damale et al., 2014; Vedani, Dobler, Dollinger, et al., 2005). By adding another dimension to 6D-QSA, a new paradigm known as 7D-QSAR is developed by including the real or virtual target-based receptor models (Polanski, 2009).

There are numerous examples based on newly developed multi-dimensional QSAR analyses. In a study by de Melo and Ferreira, inhibitory activity for 85 HIV-1 integrase strand transferase was investigated using LQTA-QSAR method. In this work, a 4D-QSAR model was generated based on conformational ensemble profile using GROMACS molecular dynamic package. The resulted model showed good predictive ability and could be used for designing novel HIV-1 integrase strand transfer inhibitors (de Melo & Ferreira, 2012). In another study, Vera-DiVaio and co-workers, developed a successful 4D-QSAR model to predict the toxicity of a new series of N-phenylpyrazole benzylidene-carbohydrazide with R^2 and Q^2 values of 0.85 and 0.79, respectively (Vera-Divaio et al., 2009). A receptor-modeling study based on 5D-QSAR analysis was performed by Vedani et al. In this study, novel designed ligands were proposed to inhibit chemokine receptor-3 (CCR3) at the low-nanomolar range using induced fit scenarios (Vedani, Dobler, & Lill, 2005). Dobler et al. developed a virtual laboratory on the Internet for the toxicity prediction of drugs on five receptor types (i.e., aryl hydrocarbon (Ah), $5HT_{2A}$, cannabinoid, $GABA_A$, and estrogen receptor) using a 5D-QSAR method (Dobler, Lill, & Vedani, 2003).

In an investigation by Peristera et al, binding mode and affinity of anabolic steroids toward mineralocorticoid receptor was studied using 6D-QSAR (Peristera, Spreafico, Smiesko, Ernst, & Vedani, 2009). The generated model was resulted from flexible docking of inhibitors to the crystal structure of the receptor with cross validated and predictive R^2 of 0.81 and 0.66, respectively. The model was claimed useful

Table 1. Dimensionality of QSAR approaches (Polanski, 2009; Vedani, Dobler, & Lill, 2006)

Dimension	Descriptors
0D-QSAR	Atom and molecular counts, molecular weight, sum of atomic properties
1D-QSAR	Fragment counts such as electronic constraints, hydrophobic constraints and steric constraints
2D-QSAR	Topological descriptors such as constitutional, topology, total polar surface area, electrostatic and quantum chemical, geometrical and molecular fingerprints property of the chemical compound
3D-QSAR	Geometrical, atomic coordinates, or energy grid descriptors and static ligand representation
4D-QSAR	Like 3D-QSAR + multiple 3D ligand representations (ensemble of conformers, orientations, protonation states, tautomers and stereoisomers)
5D-QSAR	Like 4D-QSAR + induced- fit
6D-QSAR	Like 5D-QSAR + solvation scenarios
7D-QSAR	Real receptor or virtual target-based receptor model

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

for *in silico* identification of compounds with endocrine-disrupting potential. The similar examples can be found in literature (Ducki, Mackenzie, Lawrence, & Snyder, 2005; Oberdorf, Schmidt, & Wunsch, 2010; Vedani et al., 2007; Vedani, Dobler, & Lill, 2005).

Whether or not increasing the dimensionality in QSAR studies will lead to generation of highly predictive QSAR models is a debatable issue. There are examples in which the 2D-QSAR can perform as good or better than 3D-QSAR methods in predicting the biological activity (Dastmalchi, Hamzeh-Mivehroud, & Asadpour-Zeynali, 2012; Jimenez Villalobos, Gaitan Ibarra, & Montalvo Acosta, 2013). In contrast, there are also evidences that show the increasing the dimensionality improves the predictive power of QSAR models (Katritzkya, Slavova, Dobcheva, & Karelson, 2007; Vedani, Dobler, & Lill, 2005). Following two individual studies present further examples on the performance evaluation of different multidimensional QSAR analyses. The inhibitory activities of 36 indole amide hydroxamic acids on histone deacetylases were analyzed based on 2D- (with over 800 descriptors) and 3D-QSAR (using steric and electrostatic fields) methods (Katritzkya et al., 2007), and the overall results (Table 2) showed that the performance of the QSAR model depends mostly on the type of the chosen variable selection method and the descriptors. It may be hard to predict which specific QSAR method from dimensionality point of view will perform the best. However, choosing the best method for a QSAR study can be achieved by trial-and-error or surveying the available literature. Damale et al. carried out 3D to 6D QSAR analyses using 106 estrogen receptor ligands, and concluded that the cross-validated and predictive R^2 values increase as the dimensionality of the QSAR method increases (Damale et al., 2014). Additionally, one may conclude that if the relevant method is selected, the best QSAR model can be achieved by increasing the dimension (Myint & Xie, 2010). The multi-dimensional methods (i.e., 4D-6D approaches) are new in QSAR studies and have been used just in a limited number of investigations. Therefore, it is too early to conclude that "the higher the dimensionality, the higher the performance".

Table 2. Comparison of multidimensional QSAR studies carried out on two databases by Katritzkya et al. (Katritzkya et al., 2007) and Damale et al. (Damale et al., 2014)

Study	Database	Method	R^2	R^2_{CV}	R^2_{pred}	Q^2	R^2_{ext}
2D-QSAR	36	BMLR	0.811	0.734	0.737	-	-
3D-QSAR (steric fields)	36	PLS	0.937	-	0.811	0.624	0.715
3D-QSAR (electrostatic fields)	36	PLS	0.789	-	0.642	0.280	-
3D-QSAR (electrostatic fields)	36	Weighted-LS	0.931	-	0.847	0.434	0.522
3D-QSAR (force fields)	106	AMBER(Macro- Model AND AMSOL software)	-	0.821	0.563	-	-
4D-QSAR	106	Quasar	-	0.810	0.788	-	-
5D-QSAR	106	Quasar + induced fit	-	0.872	0.790	-	-
6D-QSAR	106	Quasar + solvation	-	0.903	0.885	-	-

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

CASE STUDIES

Today, modern computational techniques revolutionized the drug design and discovery process by accelerating and economizing the procedure. There are successful QSAR-driven examples of drug design projects which yielded pharmaceuticals already in the market or under investigation. Especially during past two decades, the toxicological based QSAR studies denoted by quantitative structure toxicity relationship (QSTR) have received lots of attention (Zhu, 2013) deduced from the huge volumes of reviews and research papers in the literature carried on several endpoints such as human health risk effects including carcinogenicity, mutagenicity, liver effects and cardiac toxicity (Toropov, et al., 2014). Few applications of 2D- and 3D-QSAR methods in drug design and development processes as well as toxicity related examples have been presented in this section.

Norfloxacin (Noroxin)® is one of earliest examples of QSAR guided drug design. Koga and co-workers investigated oxoquinoline-3-carboxylic acids derivatives as the antibacterial agents and established a QSAR model in which antibacterial activity was parabolically correlated to steric parameters (Koga, Itoh, Murayama, Suzue, & Irikura, 1980). By structural modification of the compounds and subsequent *in vitro* and *in vivo* studies, Norfloxacin reached the market. This drug is used in urinary tract infections.

In an attempt to design safe anti ulcer H_2 blockers, Cimetidine (Tagamet)® was developed from metiamide through the quantitative comparison of physicochemical properties and bioisosteric replacements of thiourea group of metiamide by cyanoguanidine group (Durant et al., 1977; Topliss, 1993). When Cimetidine was marketed in 1976, it was one of the largest selling pharmaceutical products in the world.

GRID based 3D-QSAR study on inhibitors of sialidases from both influenza viruses A and B based on crystal structure of the enzyme leads to discovery of Zanamivir (Relenza)® (von Itzstein et al., 1993). Active site structure was computationally probed using GRID program and the analyses suggested energetically favorable substitutions for modification of parent structure. It is the best example of structure-based drug design approach.

In a QSAR-directed design, a data set of tylophorine derivatives as anticancer agents was studied using validated QSAR modeling and virtual screening. The original data set was split into training and test set. The variable selection *k*-nearest neighbors (*k*NNs) method was used for descriptor selection followed by statistical evaluation procedures. Among the compounds predicted to be highly active (the hits) based on QSAR analysis, the experimental studies showed that eighty percent of them were active (Zhang et al., 2007). A similar QSAR study was performed for anticonvulsant agents indicating a rational drug design of novel compounds before synthesis (Shen et al., 2004).

In our previous investigation, a 2D-QSAR study was conducted on arylbenzofuran derivatives as H_3 receptor antagonists. In this study, MLR-based QSAR models were generated using GA-PLS descriptor selection strategy with different kinds of descriptors. The predictive power of the models was assessed by applying cross validation, external validation, and Y-scrambling methods. The structural requirements needed for the antagonistic activity was identified. The result of this investigation may aid to the development of new potent H_3 antagonists (Dastmalchi, et al., 2008).

As mentioned earlier, one of the important fields of QSAR application is the prediction of toxicity behavior of drugs and hazardous materials. Numerous examples can be found in the literature on this topic and the full coverage is beyond this chapter. However, few examples are provided to address some recent cases.

The developmental toxicity is an important endpoint for risk assessments of chemical exposure in human health, and is often considered as hard to find data because of the cost and associated complexity.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

In an interesting study, Craig et al. (Craig, Wang, & Zhao, 2013) have used a set of free software and publically available toxicity databases to develop a single variable QSAR model using log of octanol–water partition coefficient (log K_{OW}) for the prediction of developmental toxicity of halogenated azole compounds. In order to increase the predictive power of the model, they elegantly extracted 35 structurally similar toxic compounds for their QSAR study, and by doing so, the developed model performs well on external dataset compounds provided that they fall within the defined applicability domain (AD).

In an investigation conducted by Zarei et al. (Zarei, Atabati, & Kor, 2014), they have used a QSAR approach based on bee colony algorithm and Adaptive neuro-fuzzy interface system (ANFIS) to predict the toxicity of 268 substituted benzene derivatives on *Tetrahymena pyriformis*. The bee algorithm was used as feature selection tool to choose three structural variables describing hydrophobicity, number of electronegative groups and the molecular shape of the studied compounds, which then they were used in the prediction step carried out by ANFIS method. As there is a good agreement between fish (aquatic animal) toxicity and toxic potency of *T. pyriformis* (Netzeva, Pavan, & Worth, 2008), the results of such *in vitro* predictions have practical applications in environmental safety assessment.

The above mentioned case studies are few examples of using QSAR methodologies, which can be used in predicting toxicity of newly designed and/or discovered drugs. However, there are numerous studies in this field which can provide useful insights into drug design, toxicity, and development processes. The data from these investigations have been organized in various toxicity databases such as OECD QSAR Toolbox[32], REACH[33], TOXNET[34], SIDER[35], ACToR[36], DailyMed[37], DSSTox[38] just to mention few (Panagiotou & Taboureau, 2012; Sullivan, et al., 2014; Toropov, et al., 2014). These databases are valuable resources for QSAR studies.

FUTURE RESEARCH DIRECTIONS

In this chapter, we have tried to introduce almost main aspects of QSAR in drug design in an stepwise manner. However, the chapter is not by any means thorough and the readers are encouraged to follow the details of different topics covered here in the supplied references. The field of QSAR studies in drug development is an open area and it seems that the current methods are far from predicting the exact activity of the test compounds. However, as a principle, the prediction methods cannot provide answer more accurate than the experimental data. The maximum accuracy one may expect to achieve from QSAR modeling is in the order of biological data. Therefore, although not an inherent feature of the modeling methods, the improved assay methods can have profound effects on the developed models. One of the main players of the QSAR studies is the way the chemicals are presented numerically by the descriptors. There are huge numbers of descriptors which have been developed to capture the molecular features. In most cases, they are calculated by programs with high speed and are used without any pre-judgment in the subsequent steps. But in most instances, the interpretation of the observed activities based on the descriptors is not obvious. Development of new methods for presenting those features of the molecular structures more relevant to drug-likeness may improve the applied aspects of the QSAR modeling efforts. The choice of which algorithm should be used for descriptor selection is also a challenging stage making it a vitally fundamental phenomena for generating QSAR models. The better selection of descriptors, the more reliable the QSAR models. The important parts of every QSAR model generation are the robustness of model prediction and the validity of the generated model. So comparing the results of several types of descriptors through multiple machine learning methods along

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

with cross validating the generated models and testing the data not included in the training set (i.e., the external validation) is an important step for QSAR experiments. The QSAR modeling problems can be related to model prediction and optimization whether on descriptors selection or model generation. Many optimization algorithms have been proposed and employed in computational modeling of drug design. The GA, PSO, ACO, ANN, and variations of them are the most applied algorithms in QSAR studies among which ANN and GA have a strong historical background in pharmaceutical drug design. The ANN as an inspiration of human brain functionality is mostly utilized in model predictions which take effects from many parameters such as feature selection, network topology design, network optimization and validation. From the recent QSAR studies, it can also be deduced that hybrid and ensemble versions of machine learning algorithms as well as graph based machine learning technologies had promising results in comparison with the original artificial intelligence algorithms. However, using the combination of the methods, whether it is in hybrid or ensemble mode, will need a higher computation power to make it possible for huge data sets of pharmaceutical drug industry to be processed. The improved computational algorithms with high efficiency and performance may aid to farther the success rate of QSAR based drug development. The last but not the least, the QSAR models can be improved by utilizing advanced hardware and computational resources.

CONCLUSION

Pharmaceutical companies always follow the cost effectiveness strategy for drug design and development. Undoubtedly, computational (*in silico*) techniques have paved the way by streamlining a number of processes for today's lead generation and identification needs. In this framework, QSAR modeling has been proved as one of the important and viable computational tools. In its widest concept, it includes finding the systematic structural differences between compounds in terms of molecular descriptors which lead to differences in corresponding biological activities. Moreover, this technique provides synthetic guidance for rational drug design prior to synthesis. In this way, the molecules possessing suitable features can be identified as hits. Over the last 50 years, QSAR studies still play an important role in identification of new chemical entities and also help to find out the borders of distinct properties of compounds for better activity. The success of QSAR models is greatly affected by accuracy of biological data obtained from experimental assays. However, gathering experimental data is not always error-free. It is important to keep in mind that, no QSAR model replaces the experimental studies and the real validation strategy is to synthesize and evaluate the new analogs. It should be emphasized that although trial-and-error factor or serendipity in drug development cannot be ruled out completely, nevertheless, QSAR and other computational approaches reduce the number of compounds needed to be synthesized and tested by identifying the most promising candidates. Although, the introduced new chemical entities do not necessarily enter the clinical trials because of the complexity of biological processes influenced by several factors such as pharmacokinetics, and lots of evaluation processes are required before reaching the market, however, the QSAR methods increase the success rate of drug development projects. The prediction of the toxicities for the chemical agents is a more complicated issue relative to that of the pharmacological effects due to limited toxicity data and related resources. However, improved *in vitro* high throughput toxicity screening test and "omics" technologies provides a roadmap for advances in QSAR-aided toxicity prediction.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

REFERENCES

Abbasi, E., Ghatee, M., & Shiri, M. E. (2013). FRAN and RBF-PSO as two components of a hyper framework to recognize protein folds. *Computers in Biology and Medicine*, *43*(9), 1182–1191. doi:10.1016/j.compbiomed.2013.05.017 PMID:23930812

Adhikari, N., Maiti, M. K., & Jha, T. (2010). Predictive comparative QSAR modelling of (phenylpiperazinyl-alkyl) oxindoles as selective 5-HT1A antagonists by stepwise regression, PCRA, FA-MLR and PLS techniques. *European Journal of Medicinal Chemistry*, *45*(3), 1119–1127. doi:10.1016/j.ejmech.2009.12.011 PMID:20053486

Afantitis, A., Melagraki, G., Sarimveis, H., Koutentis, P. A., Markopoulos, J., & Igglessi-Markopoulou, O. (2006). A novel simple QSAR model for the prediction of anti-HIV activity using multiple linear regression analysis. *Molecular Diversity*, *10*(3), 405–414. doi:10.1007/s11030-005-9012-2 PMID:16896545

Aguiar-Pulido, V., Gestal, M., Cruz-Monteagudo, M., Rabunal, J. R., Dorado, J., & Munteanu, C. R. (2013). Evolutionary computation and QSAR research. *Current Compuer Aided Drug Design*, *9*(2), 206–225. doi:10.2174/1573409911309020006 PMID:23700999

Akamatsu, M. (2002). Current state and perspectives of 3D-QSAR. *Current Topics in Medicinal Chemistry*, *2*(12), 1381–1394. doi:10.2174/1568026023392887 PMID:12470286

Andrade, C. H., Pasqualoto, K. F., Ferreira, E. I., & Hopfinger, A. J. (2010). 4D-QSAR: Perspectives in drug design. *Molecules (Basel, Switzerland)*, *15*(5), 3281–3294. doi:10.3390/molecules15053281 PMID:20657478

Angulo, J. C., & Antolin, J. (2007). Atomic quantum similarity indices in position and momentum spaces. *The Journal of Chemical Physics*, *126*(4), 044106. doi:10.1063/1.2429064 PMID:17286461

Bababadani, B. M., & Mousavi, M. (2013). Gravitational search algorithm: A new feature selection method for QSAR study of anticancer potency of imidazo[4,5-b]pyridine deravatives. *Chemometrics and Intelligent Laboratory Systems*, *122*, 1–11. doi:10.1016/j.chemolab.2012.12.002

Bagheri, S., Omidikia, N., & Kompany-Zareh, M. (2013). Unsupervised selection of inforamtive descriptors in QSAR study of anti-HIV activities of HEPT derivatives. *Chemometrics and Intelligent Laboratory Systems*, *128*, 135–143. doi:10.1016/j.chemolab.2013.08.004

Barker, M., & Rayens, W. (2003). Partial least squares for descrimination. *Journal of Chemometrics*, *17*(3), 166–173. doi:10.1002/cem.785

Baroni, M., Costantino, G., Cruciani, G., Riganelli, D., Valigi, R., & Clementi, S. (1993). Generating optimal linear PLS estimations (GOLPE): An advanced chemometric tool for handling 3D-QSAR problems. *Quantitative Structure-Activity Relationships*, *12*(1), 9–20. doi:10.1002/qsar.19930120103

Baskin, I. I., Ait, A. O., Halberstam, N. M., Palyulin, V. A., & Zefirov, N. S. (2002). An approach to the interpretation of backpropagation neural network models in QSAR studies. *SAR and QSAR in Environmental Research*, *13*(1), 35–41. doi:10.1080/10629360290002073 PMID:12074390

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Baskin, I. I., & Zhokhova, N. I. (2013). The continuous molecular fields approach to building 3D-QSAR models. *Journal of Computer-Aided Molecular Design*, *27*(5), 427–442. doi:10.1007/s10822-013-9656-4 PMID:23719959

Bonabeau, E., Dorigo, M., & Theraulaz, G. (2000). Inspiration for optimization from social insect behaviour. *Nature*, *406*(6791), 39–42. doi:10.1038/35017500 PMID:10894532

Butkiewicz, M., Lowe, E. W. Jr, Mueller, R., Mendenhall, J. L., Teixeira, P. L., Weaver, C. D., & Meiler, J. (2013). Benchmarking ligand-based virtual High-Throughput Screening with the PubChem database. *Molecules (Basel, Switzerland)*, *18*(1), 735–756. doi:10.3390/molecules18010735 PMID:23299552

Chen, S. A., Ou, Y. Y., Lee, T. Y., & Gromiha, M. M. (2011). Prediction of transporter targets using efficient RBF networks with PSSM profiles and biochemical properties. *Bioinformatics (Oxford, England)*, *27*(15), 2062–2067. doi:10.1093/bioinformatics/btr340 PMID:21653515

Chen, Y., Xu, J., Yang, B., Zhao, Y., & He, W. (2012). A novel method for prediction of protein interaction sites based on integrated RBF neural networks. *Computers in Biology and Medicine*, *42*(4), 402–407. doi:10.1016/j.compbiomed.2011.12.007 PMID:22226645

Cheng, Z., Zhang, Y., & Zhou, C. (2011). QSAR models for phosphoramidate prodrugs of 2'-methylcytidine as inhibitors of hepatitis C virus based on PSO boosting. *Chemical Biology & Drug Design*, *78*(6), 948–959. doi:10.1111/j.1747-0285.2011.01236.x PMID:21895985

Chupakhin, V., Marcou, G., Baskin, I., Varnek, A., & Rognan, D. (2013). Predicting ligand binding modes from neural networks trained on protein-ligand interaction fingerprints. *Journal of Chemical Information and Modeling*, *53*(4), 763–772. doi:10.1021/ci300200r PMID:23480697

Craig, E. A., Wang, N. C., & Zhao, Q. J. (2013). Using quantitative structure-activity relationship modeling to quantitatively predict the developmental toxicity of halogenated azole compounds. *Journal of Applied Toxicology*. doi:10.1002/jat.2940 PMID:24122872

Cramer, R. D., Patterson, D. E., & Bunce, J. D. (1988). Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *Journal of the American Chemical Society*, *110*(18), 5959–5967. doi:10.1021/ja00226a005 PMID:22148765

Cruciani, G. (Ed.). (2006). *Molecular interaction fields. Application in drug discovery and ADME prediction* (Vol. 27). Darmstadt: WILEY-VCH.

Damale, M. G., Harke, S. N., Kalam Khan, F. A., Shinde, D. B., & Sangshetti, J. N. (2014). Recent advances in multidimensional QSAR (4D-6D): A critical review. *Mini Reviews in Medicinal Chemistry*, *14*(1), 35–55. doi:10.2174/1389557513136660104 PMID:24195665

Dastmalchi, S., Hamzeh-Mivehroud, M., & Asadpour-Zeynali, K. (2012). Comparison of different 2D and 3D-QSAR methods on activity prediction of histamine H3 receptor antagonists. *Iranian Journal of Pharmaceutical Research*, *11*(1), 97–108. PMID:25317190

Dastmalchi, S., Hamzeh-Mivehroud, M., Ghafourian, T., & Hamzeiy, H. (2008). Molecular modeling of histamine H3 receptor and QSAR studies on arylbenzofuran derived H3 antagonists. *Journal of Molecular Graphics & Modelling*, *26*(5), 834–844. doi:10.1016/j.jmgm.2007.05.002 PMID:17561422

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Datar, P. A., Khedkar, S. A., Malde, A. K., & Coutinho, E. C. (2006). Comparative residue interaction analysis (CoRIA): A 3D-QSAR approach to explore the binding contributions of active site residues with ligands. *Journal of Computer-Aided Molecular Design*, *20*(6), 343–360. doi:10.1007/s10822-006-9051-5 PMID:17009094

de Melo, E. B., & Ferreira, M. M. (2012). Four-dimensional structure-activity relationship model to predict HIV-1 integrase strand transfer inhibition using LQTA-QSAR methodology. *Journal of Chemical Information and Modeling*, *52*(7), 1722–1732. doi:10.1021/ci300039a PMID:22657398

Dearden, J. C., Cronin, M. T., & Kaiser, K. L. (2009). How not to develop a quantitative structure-activity or structure-property relationship (QSAR/QSPR). *SAR and QSAR in Environmental Research*, *20*(3-4), 241–266. doi:10.1080/10629360902949567 PMID:19544191

Dobler, M., Lill, M. A., & Vedani, A. (2003). From crystal structures and their analysis to the in silico prediction of toxic phenomena1. *Helvetica Chimica Acta*, *86*(5), 1554–1568. doi:10.1002/hlca.200390134

Doo Ho, C., Sung Kwang, L., Bum Tae, K., & Kyoung Tai, N. (2001). Quantitative structure-activity relationship (QSAR) study of new fluorovinyloxyacetamides. *Bulletin of the Korean Chemical Society*, *22*(4), 388–394.

Dorigo, M., Maniezzo, V., & Colorni, A. (1996). Ant system: Optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man, and Cybernetics*, *26*(1), 29–41. doi:10.1109/3477.484436 PMID:18263004

Du, H., Wang, J., Hu, Z., & Yao, X. (2008). Quantitative structure-retention relationship study of the constituents of saffron aroma in SPME-GC-MS based on the projection pursuit regression method. *Talanta*, *77*(1), 360–365. doi:10.1016/j.talanta.2008.06.038 PMID:18804646

Du, H., Zhang, X., Wang, J., Yao, X., & Hu, Z. (2008). Novel approaches to predict the retention of histidine-containing peptides in immobilized metal-affinity chromatography. *Proteomics*, *8*(11), 2185–2195. doi:10.1002/pmic.200700788 PMID:18446801

Du, Q. S., Huang, R. B., Wei, Y. T., Pang, Z. W., Du, L. Q., & Chou, K. C. (2009). Fragment-based quantitative structure-activity relationship (FB-QSAR) for fragment-based drug design. *Journal of Computational Chemistry*, *30*(2), 295–304. doi:10.1002/jcc.21056 PMID:18613071

Ducki, S., Mackenzie, G., Lawrence, N. J., & Snyder, J. P. (2005). Quantitative structure-activity relationship (5D-QSAR) study of combretastatin-like analogues as inhibitors of tubulin assembly. *Journal of Medicinal Chemistry*, *48*(2), 457–465. doi:10.1021/jm049444m PMID:15658859

Dudek, A. Z., Arodz, T., & Galvez, J. (2006). Computational methods in developing quantitative structure-activity relationships (QSAR): A review. *Combinatorial Chemistry & High Throughput Screening*, *9*(3), 213–228. doi:10.2174/138620706776055539 PMID:16533155

Duran, A., Zamora, I., & Pastor, M. (2009). Suitability of GRIND-based principal properties for the description of molecular similarity and ligand-based virtual screening. *Journal of Chemical Information and Modeling*, *49*(9), 2129–2138. doi:10.1021/ci900228x PMID:19728739

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Durant, G. J., Emmett, J. C., Ganellin, C. R., Miles, P. D., Parsons, M. E., Prain, H. D., & White, G. R. (1977). Cyanoguanidine-thiourea equivalence in the development of the histamine H2-receptor antagonist, cimetidine. *Journal of Medicinal Chemistry*, *20*(7), 901–906. doi:10.1021/jm00217a007 PMID:17751

Dury, L., Latour, T., Leherte, L., Barberis, F., & Vercauteren, D. P. (2001). A new graph descriptor for molecules containing cycles. Application as screening criterion for searching molecular structures within large databases of organic compounds. *J Journal of Chemical Information and Computer Sciences*, *41*(6), 1437–1445. PMID:11749566

Dutt, R., & Madan, A. K. (2012). Predicting biological activity: Computational approach using novel distance based molecular descriptors. *Computers in Biology and Medicine*, *42*(10), 1026–1041. doi:10.1016/j.compbiomed.2012.08.006 PMID:22964398

El Kerdawy, A., Gussregen, S., Matter, H., Hennemann, M., & Clark, T. (2013). Quantum mechanics-based properties for 3D-QSAR. *Journal of Chemical Information and Modeling*, *53*(6), 1486–1502. doi:10.1021/ci400181b PMID:23692495

Esposito Vinzi, V. (2010). *Handbook of partial least squares: concepts, methods and applications*. Berlin: Springer. doi:10.1007/978-3-540-32827-8

Estrada, E., & Molina, E. (2001). Novel local (fragment-based) topological molecular descriptors for QSpr/QSAR and molecular design. *Journal of Molecular Graphics & Modelling*, *20*(1), 54–64. doi:10.1016/S1093-3263(01)00100-0 PMID:11760003

Farahani, M. R. (2013). Third-connectivity and third-sum-connectivity indices of circumcoronene series of benzenoid H(k). *Acta Chimica Slovenica*, *60*(1), 198–202. PMID:23841353

Fernandez, M., Caballero, J., Fernandez, L., & Sarai, A. (2011). Genetic algorithm optimization in drug design QSAR: Bayesian-regularized genetic neural networks (BRGNN) and genetic algorithm-optimized support vectors machines (GA-SVM). *Molecular Diversity*, *15*(1), 269–289. doi:10.1007/s11030-010-9234-9 PMID:20306130

Fernandez-Blanco, E., Aguiar-Pulido, V., Munteanu, C. R., & Dorado, J. (2013). Random Forest classification based on star graph topological indices for antioxidant proteins. *Journal of Theoretical Biology*, *317*, 331–337. doi:10.1016/j.jtbi.2012.10.006 PMID:23116665

Fernandez-Lozano, C., Gestal, M., Pedreira-Souto, N., Postelnicu, L., Dorado, J., & Munteanu, C. R. (2013). Kernel-based feature selection techniques for transport proteins based on star graph topological indices. *Current Topics in Medicinal Chemistry*, *13*(14), 1681–1691. doi:10.2174/15680266113139990119 PMID:23889046

Ferraty, F. r., & Romain, Y. (2011). *The Oxford handbook of functional data analysis*. Oxford, UK: Oxford University Press.

Ghosh, P., & Bagchi, M. C. (2009). QSAR modeling for quinoxaline derivatives using genetic algorithm and simulated annealing based feature selection. *Current Medicinal Chemistry*, *16*(30), 4032–4048. doi:10.2174/092986709789352303 PMID:19747124

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Gohlke, H., & Klebe, G. (2002). DrugScore meets CoMFA: Adaptation of fields for molecular comparison (AFMoC) or how to tailor knowledge-based pair-potentials to a particular protein. *Journal of Medicinal Chemistry*, *45*(19), 4153–4170. doi:10.1021/jm020808p PMID:12213058

Golbraikh, A., & Tropsha, A. (2002). Beware of q2! *Journal of Molecular Graphics & Modelling*, *20*(4), 269–276. doi:10.1016/S1093-3263(01)00123-1 PMID:11858635

Goldberg, D. E. (1989). *Genetic algorithms in search, optimization and machine learning*. Reading, MA: Addison Welsey.

Gonzalez, M. P., Teran, C., Saiz-Urra, L., & Teijeira, M. (2008). Variable Selection Methods in QSAR: An overview. *Current Topics in Medicinal Chemistry*, *8*(18), 1606–1627. doi:10.2174/156802608786786552 PMID:19075770

Goodarzi, M., & Freitas, M. P. (2010). MIA-QSAR coupled to principal component analysis-adaptive neuro-fuzzy inference systems (PCA-ANFIS) for the modeling of the anti-HIV reverse transcriptase activities of TIBO derivatives. *European Journal of Medicinal Chemistry*, *45*(4), 1352–1358. doi:10.1016/j.ejmech.2009.12.028 PMID:20060625

Goodarzi, M., Jensen, R., & Vander Heyden, Y. (2012). QSRR modeling for diverse drugs using different feature selection methods coupled with linear and nonlinear regressions. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences*, *910*, 84–94. doi:10.1016/j.jchromb.2012.01.012 PMID:22341354

Goodarzi, M., Saeys, W., Araujo, M. C., Galvao, R. K., & Vander Heyden, Y. (2013). Binary classification of chalcone derivatives with LDA or KNN based on their antileishmanial activity and molecular descriptors selected using the Successive Projections Algorithm feature-selection technique. *European Journal of Pharmaceutical Sciences*. doi:10.1016/j.ejps.2013.09.019 PMID:24090733

Goodarzi, M., Saeys, W., Deeb, O., Pieters, S., & Vander Heyden, Y. (2013). Particle swarm optimization and genetic algorithm as feature selection techniques for the QSAR modeling of imidazo[1,5-a]pyrido[3,2-e]pyrazines, inhibitors of phosphodiesterase 10A. *Chemical Biology & Drug Design*, *82*(6), 685–696. doi:10.1111/cbdd.12196 PMID:23906083

Goodford, P. J. (1985). A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *Journal of Medicinal Chemistry*, *28*(7), 849–857. doi:10.1021/jm00145a002 PMID:3892003

Goodsell, D. S. (2009). Computational docking of biomolecular complexes with AutoDock. *Cold Spring Harbor Protocols, 2009*(5). doi: 10.1101/pdb.prot5200

Goodsell, D. S., Morris, G. M., & Olson, A. J. (1996). Automated docking of flexible ligands: Applications of AutoDock. *Journal of Molecular Recognition*, *9*(1), 1–5. doi:10.1002/(SICI)1099-1352(199601)9:1<1::AID-JMR241>3.0.CO;2-6 PMID:8723313

Gozalbes, R., Doucet, J. P., & Derouin, F. (2002). Application of topological descriptors in QSAR and drug design: History and new trends. *Current Drug Targets. Infectious Disorders*, *2*(1), 93–102. doi:10.2174/1568005024605909 PMID:12462157

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Gramatica, P. (2007). Principles of QSAR models validation: Internal and external. *QSAR & Combinatorial Science*, *26*(5), 694–701. doi:10.1002/qsar.200610151

Gussregen, S., Matter, H., Hessler, G., Muller, M., Schmidt, F., & Clark, T. (2012). 3D-QSAR based on quantum-chemical molecular fields: Toward an improved description of halogen interactions. *Journal of Chemical Information and Modeling*, *52*(9), 2441–2453. doi:10.1021/ci300253z PMID:22917472

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, *46*(1/3), 389–422. doi:10.1023/A:1012487302797

Hall, L. H., & Kier, L. B. (1991). The molecular connectivity chi indexes and kappa shape indexes in structure-property modeling. In K. B. Lipkowitz & D. B. Boyd (Eds.), *Reviews in computational chemistry* (pp. 367–422). New York: VCH. doi:10.1002/9780470125793.ch9

Hall, L. H., & Kier, L. B. (2001). Issues in representation of molecular structure the development of molecular connectivity. *Journal of Molecular Graphics & Modelling*, *20*(1), 4–18. doi:10.1016/S1093-3263(01)00097-3 PMID:11760002

Hammett, L. P. (1937). The effect of structure upon the reactions of organic compounds: Benzene derivatives. *Journal of the American Chemical Society*, *59*(1), 96–103. doi:10.1021/ja01280a022

Hansch, C., Maloney, P. P., Fujita, T., & Muir, R. M. (1962). Correlation of biological activity of phenoxyacetic acids with hammett substituent constants and partition coefficients. *Nature*, *194*(4824), 178–180. doi:10.1038/194178b0

Hardin, C. T., & Rouchka, E. C. (2005). DNA motif detection using particle swarm optimization and expectation-maximization.*IEEE Symposium on Swarm Intelligence*, 2005, 181-184. doi:10.1109/SIS.2005.1501620

He, Y., Chen, D., & Zhao, W. (2006). Ensemble classifier syste based on ant colony algorithm and its application in chemical pattern classification. *Chemometrics and Intelligent Laboratory Systems*, *80*(1), 39–49. doi:10.1016/j.chemolab.2005.06.003

Hemmateenejad, B., Yousefinejad, S., & Mehdipour, A. R. (2011). Novel amino acids indices based on quantum topological molecular similarity and their application to QSAR study of peptides. *Amino Acids*, *40*(4), 1169–1183. doi:10.1007/s00726-010-0741-x PMID:20852906

Higo, J. I., & GÅ, N. (. (1989). Algorithm for rapid calculation of excluded volume of large molecules. *Journal of Computational Chemistry*, *10*(3), 376–379. doi:10.1002/jcc.540100311

Holland, J. (1975). *Adaptation in natrual and artificial systems*. Ann Arbor: University of Michigan Press.

Hopfinger, A. J., Wang, S., Tokarski, J. S., Jin, B., Albuquerque, M., Madhav, P. J., & Duraiswami, C. (1997). Construction of 3D-QSAR models using the 4D-QSAR analysis formalism. *Journal of the American Chemical Society*, *119*(43), 10509–10524. doi:10.1021/ja9718937

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology, 24*, 417-441, 498-520.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Ivanciuc, O. (1999). The neural network MolNet prediction of alkane enthalpies. *Analytica Chimica Acta*, *384*(3), 271–284. doi:10.1016/S0003-2670(98)00777-6

Ivanciuc, O. (2003). Graph theory in chemistry. In J. Gasteiger (Ed.), *Handbook of cheminformatics* (Vol. 1, pp. 103–138). Weinheim: Wiley-VCH. doi:10.1002/9783527618279.ch6

Ivanciuc, O. (2013). Chemical graphs, molecular matrices and topological indices in cheminformatics and quantitative structure-activity relationships. *Current Computer-aided Drug Design*, *9*(2), 153–163. doi:10.2174/1573409911309020002 PMID:23701000

Izrailev, S., & Agrafiotis, D. (2001). A novel method for building regression tree models for QSAR based on artificial ant colony systems. *Journal of Chemical Information and Computer Sciences*, *41*(1), 176–180. doi:10.1021/ci000336s PMID:11206370

Izrailev, S., & Agrafiotis, D. K. (2002). Variable selection for QSAR by artificial ant colony systems. *SAR and QSAR in Environmental Research*, *13*(3-4), 417–423. doi:10.1080/10629360290014296 PMID:12184383

Jackson, C., Esnouf, M. P., Winzor, D., & Duewer, D. (2007). Defining and measuring biological activity: Applying the principles of metrology. *Accreditation and Quality Assurance*, *12*(6), 283–294. doi:10.1007/s00769-006-0254-1

Jalali-Heravi, M., & Asadollahi-Baboli, M. (2009). Quantitative structure-activity relationship study of serotonin (5-HT7) receptor inhibitors using modified ant colony algorithm and adaptive neuro-fuzzy interference system (ANFIS). *European Journal of Medicinal Chemistry*, *44*(4), 1463–1470. doi:10.1016/j.ejmech.2008.09.050 PMID:19013691

Jimenez Villalobos, T. P., Gaitan Ibarra, R., & Montalvo Acosta, J. J. (2013). 2D, 3D-QSAR and molecular docking of 4(1H)-quinolones analogues with antimalarial activities. *Journal of Molecular Graphics & Modelling*, *46*, 105–124. doi:10.1016/j.jmgm.2013.10.002 PMID:24185373

Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). New York: Springer.

Jores-Kong, H., & Wold, H. (1982). *Systems under indirect observation: Causality, structure, prediction*. Amsterdam, The Netherlands: North-Holland.

Kadam, R. U., & Roy, N. (2006). Cluster analysis and two-dimensional quantitative structure-activity relationship (2D-QSAR) of Pseudomonas aeruginosa deacetylase LpxC inhibitors. *Bioorganic & Medicinal Chemistry Letters*, *16*(19), 5136–5143. doi:10.1016/j.bmcl.2006.07.041 PMID:16879960

Karelson, M., Lobanov, V. S., & Katritzky, A. R. (1996). Quantum-chemical descriptors in QSAR/QSPR studies. *Chemical Reviews*, *96*(3), 1027–1044. doi:10.1021/cr950202r PMID:11848779

Katritzky, A. R., Dobchev, D. A., Fara, D. C., Hur, E., Tamm, K., & Kurunczi, L. et al. (2006). Skin permeation rate as a function of chemical structure. *Journal of Medicinal Chemistry*, *49*(11), 3305–3314. doi:10.1021/jm051031d PMID:16722649

Katritzky, A. R., & Gordeeva, E. V. (1993). Traditional topological indices vs electronic, geometrical, and combined molecular descriptors in QSAR/QSPR research. *J Journal of Chemical Information and Computer Sciences*, *33*(6), 835–857. PMID:8113335

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Katritzky, A. R., Kuanar, M., Fara, D. C., Karelson, M., Acree, W. E. Jr, Solov'ev, V. P., & Varnek, A. (2005). QSAR modeling of blood:air and tissue:air partition coefficients using theoretical descriptors. *Bioorganic & Medicinal Chemistry*, *13*(23), 6450–6463. doi:10.1016/j.bmc.2005.06.066 PMID:16202613

Katritzky, A. R., Kuanar, M., Slavov, S., Dobchev, D. A., Fara, D. C., & Karelson, M. et al. (2006). Correlation of blood-brain penetration using structural descriptors. *Bioorganic & Medicinal Chemistry*, *14*(14), 4888–4917. doi:10.1016/j.bmc.2006.03.012 PMID:16697202

Katritzky, A. R., Lobanov, V. S., & Karelson, M. (1995). QSPR: The correlation and quantitative prediction of chemical and physical properties from structure. *Chemical Society Reviews*, *24*(4), 279–287. doi:10.1039/cs9952400279

Katritzky, A. R., Lobanov, V. S., & Karelson, M. (2007). *Comprehensive descriptors for structural and statistical analysis*. CODESSA.

Katritzkya, A. R., Slavova, S. H., Dobcheva, D. A., & Karelson, M. (2007). Comparison between 2D and 3D-QSAR approaches to correlate inhibitor activity for a series of indole amide hydroxamic acids. *QSAR & Combinatorial Science*, *26*(3), 333–345. doi:10.1002/qsar.200630021

Kennedy, J., & Eberhart, R. (1995). *Particle swarm optimization.* Paper presented at the IEEE International Conference on Neural Networks. doi:10.1109/ICNN.1995.488968

Kier, L. B., & Hall, L. H. (1986). *Molecular connectivity in structure-activity analysis*. New York: Wiley.

Kireeva, D. B. (1995). ChemNet: A novel neural network based method for graph/property mapping. *Journal of Chemical Information and Computer Sciences*, *35*, 175–180.

Klebe, G., Abraham, U., & Mietzner, T. (1994). Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *Journal of Medicinal Chemistry*, *37*(24), 4130–4146. doi:10.1021/jm00050a010 PMID:7990113

Koga, H., Itoh, A., Murayama, S., Suzue, S., & Irikura, T. (1980). Structure-activity relationships of antibacterial 6,7- and 7,8-disubstituted 1-alkyl-1,4-dihydro-4-oxoquinoline-3-carboxylic acids. *Journal of Medicinal Chemistry*, *23*(12), 1358–1363. doi:10.1021/jm00186a014 PMID:7452690

Kubinyi, H. (1993). *QSAR: Hansch analysis and related approaches*. Weinheim: VCH. doi:10.1002/9783527616824

Leach, A. R., & Gillet, V. J. (2007). *An introduction to cheminformatics*. Dordrecht: Springer. doi:10.1007/978-1-4020-6291-9

Leonard, J. T., & Roy, K. (2006). QSAR by LFER model of HIV protease inhibitor mannitol derivatives using FA-MLR, PCRA, and PLS techniques. *Bioorganic & Medicinal Chemistry*, *14*(4), 1039–1046. doi:10.1016/j.bmc.2005.09.022 PMID:16213730

Li, B. K., Cong, Y., Yang, X. G., Xue, Y., & Chen, Y. Z. (2013). In silico prediction of spleen tyrosine kinase inhibitors using machine learning approaches and an optimized molecular descriptor subset generated by recursive feature elimination method. *Computers in Biology and Medicine*, *43*(4), 395–404. doi:10.1016/j.compbiomed.2013.01.015 PMID:23402937

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Li, G.-Z., & Zeng, X.-Q. (2009). Feature selection for partial least square based dimension reduction. *Studies in Computational Intelligence*, *205*, 3–37. doi:10.1007/978-3-642-01536-6_1

Liew, C. Y., & Yap, C. W. (2012). Current modeling methods used in QSAR/QSPR. In M. Dehmer, K. Varmuza, & D. Bonchev (Eds.), *Statistical modelling of molcular descriptors in QSAR/QSPR* (pp. 1–31). Wiley.

Lin, T. H., Li, H. T., & Tsai, K. C. (2004). Implementing the Fisher's discriminant ratio in a k-means clustering algorithm for feature selection and data set trimming. *Journal of Chemical Information and Computer Sciences*, *44*(1), 76–87. doi:10.1021/ci030295a PMID:14741013

Liu, P., & Long, W. (2009). Current mathematical methods used in QSAR/QSPR studies. *International Journal of Molecular Sciences*, *10*(5), 1978–1998. doi:10.3390/ijms10051978 PMID:19564933

Livingstone, D. J., & Salt, D. W. (2005). Judging the significance of multiple linear regression models. *Journal of Medicinal Chemistry*, *48*(3), 661–663. doi:10.1021/jm049111p PMID:15689150

Lozano, N. B., Maltarollo, V. G., Weber, K. C., Honorio, K. M., Guido, R. V., Andricopulo, A. D., & Da Silva, A. B. (2012). Molecular features for antitrypanosomal activity of thiosemicarbazones revealed by OPS-PLS QSAR studies. *Medicinal Chemistry (Shariqah, United Arab Emirates)*, *8*(6), 1045–1056. PMID:22779790

Lu, J. X., Shen, Q., Jiang, J. H., Shen, G. L., & Yu, R. Q. (2004). QSAR analysis of cyclooxygenase inhibitor using particle swarm optimization and multiple linear regression. *Journal of Pharmaceutical and Biomedical Analysis*, *35*(4), 679–687. doi:10.1016/j.jpba.2004.02.026 PMID:15193712

Lushington, G. H., Guo, J. X., & Wang, J. L. (2007). Whither combine? New opportunities for receptor-based QSAR. *Current Medicinal Chemistry*, *14*(17), 1863–1877. doi:10.2174/092986707781058814 PMID:17627522

Maccari, L., Magnani, M., Strappaghetti, G., Corelli, F., Botta, M., & Manetti, F. (2006). A genetic-function-approximation-based QSAR model for the affinity of arylpiperazines toward alpha1 adrenoceptors. *Journal of Chemical Information and Modeling*, *46*(3), 1466–1478. doi:10.1021/ci060031z PMID:16711766

Martinez, A. M., & Kak, A. C. (2001). PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*(2), 228–233. doi:10.1109/34.908974

Martins, J. P., Barbosa, E. G., Pasqualoto, K. F., & Ferreira, M. M. (2009). LQTA-QSAR: A new 4D-QSAR methodology. *Journal of Chemical Information and Modeling*, *49*(6), 1428–1436. doi:10.1021/ci900014f PMID:19422246

Mc, C. W., & Pitts, W. (1948). The statistical organization of nervous activity. *Biometrics*, *4*(2), 91–99. doi:10.2307/3001453 PMID:18871168

Mercader, A. G., Duchowicz, P. R., Fernández, F. M., & Castro, E. A. (2008). Modified and enhanced replacement method for the selection of molecular descriptors in QSAR and QSPR theories. *Chemometrics and Intelligent Laboratory Systems*, *92*(2), 138–144. doi:10.1016/j.chemolab.2008.02.005

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Mercader, A. G., Duchowicz, P. R., Fernandez, F. M., & Castro, E. A. (2010). Replacement method and enhanced replacement method versus the genetic algorithm approach for the selection of molecular descriptors in QSPR/QSAR theories. *Journal of Chemical Information and Modeling*, *50*(9), 1542–1548. doi:10.1021/ci100103r PMID:20722426

Mercader, A. G., Duchowicz, P. R., Fernandez, F. M., & Castro, E. A. (2011). Advances in the replacement and enhanced replacement method in QSAR and QSPR theories. *Journal of Chemical Information and Modeling*, *51*(7), 1575–1581. doi:10.1021/ci200079b PMID:21644502

Moody, J., & Darken, C. (1989). Fast learning in networks of locally-tuned processing units. *Neural Computation*, *1*(2), 281–294. doi:10.1162/neco.1989.1.2.281

Morris, G. M., Goodsell, D. S., Huey, R., & Olson, A. J. (1996). Distributed automated docking of flexible ligands to proteins: Parallel applications of AutoDock 2.4. *Journal of Computer-Aided Molecular Design*, *10*(4), 293–304. doi:10.1007/BF00124499 PMID:8877701

Mozrzymas, A. (2013). Modelling of the critical micelle concentration of cationic gemini surfactants using molecular connectivity indices. *Journal of Solution Chemistry*, *42*(11), 2187–2199. doi:10.1007/s10953-013-0095-6 PMID:24319301

Munteanu, C. R., Fernandez-Blanco, E., Seoane, J. A., Izquierdo-Novo, P., Rodriguez-Fernandez, J. A., & Prieto-Gonzalez, J. M. et al. (2010). Drug discovery and design for complex diseases through QSAR computational methods. *Current Pharmaceutical Design*, *16*(24), 2640–2655. doi:10.2174/138161210792389252 PMID:20642425

Myint, K. Z., & Xie, X. Q. (2010). Recent advances in fragment-based QSAR and multi-dimensional QSAR methods. *International Journal of Molecular Sciences*, *11*(10), 3846–3866. doi:10.3390/ijms11103846 PMID:21152304

Netzeva, T. I., Pavan, M., & Worth, A. P. (2008). Review of (quantitative) structure–activity relationships for acute aquatic toxicity. *QSAR & Combinatorial Science*, *27*(1), 77–90. doi:10.1002/qsar.200710099

Noorizadeh, H., Sajjadifar, S., & Farmany, A. (2013). A quantitative structure-activity relationship study of anti-HIV activity of substituted HEPT using nonlinear models. *Medicinal Chemistry Research*, *22*(11), 5442–5452. doi:10.1007/s00044-013-0525-4 PMID:24098069

Oberdorf, C., Schmidt, T. J., & Wunsch, B. (2010). 5D-QSAR for spirocyclic sigma1 receptor ligands by Quasar receptor surface modeling. *European Journal of Medicinal Chemistry*, *45*(7), 3116–3124. doi:10.1016/j.ejmech.2010.03.048 PMID:20427100

Olah, M., Bologa, C., & Oprea, T. I. (2004). An automated PLS search for biologically relevant QSAR descriptors. *Journal of Computer-Aided Molecular Design*, *18*(7-9), 437–449. doi:10.1007/s10822-004-4060-8 PMID:15729845

Oliveira Junior, H. A. e. (2012). *Stochastic global optimization and its applications with fuzzy adaptive simulated annealing.* New York: Springer-Verlag. doi:10.1007/978-3-642-27479-4

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Ou, Y. Y., Gromiha, M. M., Chen, S. A., & Suwa, M. (2008). TMBETADISC-RBF: Discrimination of beta-barrel membrane proteins using RBF networks and PSSM profiles. *Computational Biology and Chemistry*, *32*(3), 227–231. doi:10.1016/j.compbiolchem.2008.03.002 PMID:18434251

Ou-Yang, S. S., Lu, J. Y., Kong, X. Q., Liang, Z. J., Luo, C., & Jiang, H. (2012). Computational drug discovery. *Acta Pharmacologica Sinica*, *33*(9), 1131–1140. doi:10.1038/aps.2012.109 PMID:22922346

Panagiotou, G., & Taboureau, O. (2012). The impact of network biology in pharmacology and toxicology. *SAR and QSAR in Environmental Research*, *23*(3-4), 221–235. doi:10.1080/1062936X.2012.657237 PMID:22352466

Pastor, M., Cruciani, G., McLay, I., Pickett, S., & Clementi, S. (2000). GRid-INdependent descriptors (GRIND): A novel class of alignment-independent three-dimensional molecular descriptors. *Journal of Medicinal Chemistry*, *43*(17), 3233–3243. doi:10.1021/jm000941m PMID:10966742

Patil, D., Raj, R., Shingade, P., Kulkarni, B., & Jayaraman, V. K. (2009). Feature selection and classification employing hybrid ant colony optimization/random forest methodology. *Combinatorial Chemistry & High Throughput Screening*, *12*(5), 507–513. doi:10.2174/138620709788488993 PMID:19519330

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, *2*(11), 559–572. doi:10.1080/14786440109462720

Perez-Castillo, Y., Lazar, C., Taminau, J., Froeyen, M., Cabrera-Perez, M. A., & Nowe, A. (2012). GA(M) E-QSAR: A novel, fully automatic genetic-algorithm-(meta)-ensembles approach for binary classification in ligand-based drug design. *Journal of Chemical Information and Modeling*, *52*(9), 2366–2386. doi:10.1021/ci300146h PMID:22856471

Peristera, O., Spreafico, M., Smiesko, M., Ernst, B., & Vedani, A. (2009). Mixed-model QSAR at the human mineralocorticoid receptor: Predicting binding mode and affinity of anabolic steroids. *Toxicology Letters*, *189*(3), 219–224. doi:10.1016/j.toxlet.2009.05.025 PMID:19523507

Pitts, W., & Mc, C. W. (1947). How we know universals: The perception of auditory and visual forms. *The Bulletin of Mathematical Biophysics*, *9*(3), 127–147. doi:10.1007/BF02478291 PMID:20262674

Polanski, J. (2009). Receptor dependent multidimensional QSAR for modeling drug-receptor interactions. *Current Medicinal Chemistry*, *16*(25), 3243–3257. doi:10.2174/092986709788803286 PMID:19548875

Polanski, J., Gieleciak, R., & Bak, A. (2002). The comparative molecular surface analysis (COMSA)--a nongrid 3D QSAR method by a coupled neural network and PLS system: Predicting pK(a) values of benzoic and alkanoic acids. *Journal of Chemical Information and Computer Sciences*, *42*(2), 184–191. doi:10.1021/ci010031t PMID:11911685

Polanski, J., & Walczak, B. (2000). The comparative molecular surface analysis (COMSA): A novel tool for molecular design. *Computers & Chemistry*, *24*(5), 615–625. doi:10.1016/S0097-8485(00)00064-4 PMID:10890372

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Pourbasheer, E., Riahi, S., Ganjali, M. R., & Norouzi, P. (2010). Quantitative structure-activity relationship (QSAR) study of interleukin-1 receptor associated kinase 4 (IRAK-4) inhibitor activity by the genetic algorithm and multiple linear regression (GA-MLR) method. *Journal of Enzyme Inhibition and Medicinal Chemistry*, *25*(6), 844–853. doi:10.3109/14756361003757893 PMID:20429783

Prakash, O., & Khan, F. (2013). Cluster based SVR-QSAR modelling for HTS records: An implementation for anticancer leads against human breast cancer. *Combinatorial Chemistry & High Throughput Screening*, *16*(7), 511–521. doi:10.2174/1386207311316070002 PMID:23488974

Raichurkar, A. V., Shah, U. A., & Kulkarni, V. M. (2011). 3D-QSAR of novel phosphodiesterase-4 inhibitors by genetic function approximation. *Medicinal Chemistry (Shariqah, United Arab Emirates)*, *7*(6), 543–552. doi:10.2174/157340611797928352 PMID:22313294

Randic, M. (2001). The connectivity index 25 years after. *Journal of Molecular Graphics & Modelling*, *20*(1), 19–35. doi:10.1016/S1093-3263(01)00098-5 PMID:11760000

Rekker, R. F., & Mannhold, R. (1992). *Calculation of drug lipophilicity: The hydrophobic fragmental constant approach* (Vol. 12). Weinheim: VCH.

Rosenblatt, F. (1962). Principles of neurodynamics: Perceptrons and the theory of brain mechanisms. Academic Press.

Roy, K., & Pratim Roy, P. (2009). Comparative chemometric modeling of cytochrome 3A4 inhibitory activity of structurally diverse compounds using stepwise MLR, FA-MLR, PLS, GFA, G/PLS and ANN techniques. *European Journal of Medicinal Chemistry*, *44*(7), 2913–2922. doi:10.1016/j.ejmech.2008.12.004 PMID:19128860

Sahin, K., Saripinar, E., Yanmaz, E., & Gecen, N. (2011). Quantitative bioactivity prediction and pharmacophore identification for benzotriazine derivatives using the electron conformational-genetic algorithm in QSAR. *SAR and QSAR in Environmental Research*, *22*(3), 217–238. doi:10.1080/1062936X.2010.548341 PMID:21391137

Sapre, N. S., Pancholi, N., Gupta, S., & Sapre, N. (2008). Computational modeling of tetrahydroimidazo-[4,5,1-jk][1,4]-benzodiazepinone derivatives: An atomistic drug design approach using Kier-Hall electrotopological state (E-state) indices. *Journal of Computational Chemistry*, *29*(11), 1699–1706. doi:10.1002/jcc.20931 PMID:18351618

Scior, T., Medina-Franco, J. L., Do, Q. T., Martinez-Mayorga, K., Yunes Rojas, J. A., & Bernard, P. (2009). How to recognize and workaround pitfalls in QSAR studies: A critical review. *Current Medicinal Chemistry*, *16*(32), 4297–4313. doi:10.2174/092986709789578213 PMID:19754417

Shamsipur, M., Zare-Shahabadi, V., Hemmateenejad, B., & Akhond, M. (2009). An efficient variable selection method based on the use of external memory in ant colony optimization. Application to QSAR/QSPR studies. *Analytica Chimica Acta*, *646*(1-2), 39–46. doi:10.1016/j.aca.2009.05.005 PMID:19523554

Shen, M., Beguin, C., Golbraikh, A., Stables, J. P., Kohn, H., & Tropsha, A. (2004). Application of predictive QSAR models to database mining: Identification and experimental validation of novel anticonvulsant compounds. *Journal of Medicinal Chemistry*, *47*(9), 2356–2364. doi:10.1021/jm030584q PMID:15084134

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Shen, Q., Jiang, J. H., Tao, J. C., Shen, G. L., & Yu, R. Q. (2005). Modified ant colony optimization algorithm for variable selection in QSAR modeling: QSAR studies of cyclooxygenase inhibitors. *Journal of Chemical Information and Modeling*, *45*(4), 1024–1029. doi:10.1021/ci049610z PMID:16045297

Shi, W. M., Shen, Q., Kong, W., & Ye, B. X. (2007). QSAR analysis of tyrosine kinase inhibitor using modified ant colony optimization and multiple linear regression. *European Journal of Medicinal Chemistry*, *42*(1), 81–86. doi:10.1016/j.ejmech.2006.08.001 PMID:16996653

Silverman, B. D., & Platt, D. E. (1996). Comparative molecular moment analysis (CoMMA): 3D-QSAR without molecular superposition. *Journal of Medicinal Chemistry*, *39*(11), 2129–2140. doi:10.1021/jm950589q PMID:8667357

Sullivan, K. M., Manuppello, J. R., & Willett, C. E. (2014). Building on a solid foundation: SAR and QSAR as a fundamental strategy to reduce animal testing. *SAR and QSAR in Environmental Research*, 1–9. doi:10.1080/1062936x.2014.907203 PMID:24773450

Sun, M., Zheng, Y., Wei, H., Chen, J., Cai, J., & Ji, M. (2009). Enhanced replacement method-based quantitative structure–activity relationship modeling and support vector machine classification of 4-anilino-3-quinolinecarbonitriles as src kinase inhibitors. *QSAR & Combinatorial Science*, *28*(3), 312–324. doi:10.1002/qsar.200860107

Sun, M., Zheng, Y., Wei, H., Chen, J., & Ji, M. (2009). QSAR studies on 4-anilino-3-quinolinecarbonitriles as Src kinase inhibitors using robust PCA and both linear and nonlinear models. *Journal of Enzyme Inhibition and Medicinal Chemistry*, *24*(5), 1109–1116. doi:10.1080/14756360802632906 PMID:19555174

Sun, Y., Todorovic, S., & Goodison, S. (2010). Local-learning-based feature selection for high-dimensional data analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(9), 1610–1626. doi:10.1109/TPAMI.2009.190 PMID:20634556

Taft, R. W. Jr. (1952). Polar and steric substituent constants for aliphatic and o-benzoate groups from rates of esterification and hydrolysis of esters1. *Journal of the American Chemical Society*, *74*(12), 3120–3128. doi:10.1021/ja01132a049

Tiwari, A., Saxena, S., Pant, A. B., & Srivastava, P. (2012). Protein-ligand interaction studies of retinol-binding protein 3 with herbal molecules using AutoDock for the management of Eales' disease. *Journal of Ocular Biology, Diseases, and Informatics, 5*(2), 40-43. doi: 10.1007/s12177-012-9098-6

Todeschini, R., & Consonni, V. (2000). *Handbook of molecular descriptors* (Vol. 11). Weinheim: WILEY-VCH Verlag GmbH. doi:10.1002/9783527613106

Topliss, J. (1993). Some observations on classical QSAR. *Perspectives in Drug Discovery and Design, 1*(2), 253-268.

Toropov, A. A., Toropova, A. P., Raska, I. Jr, Leszczynska, D., & Leszczynski, J. (2014). Comprehension of drug toxicity: Software and databases. *Computers in Biology and Medicine*, *45*, 20–25. doi:10.1016/j.compbiomed.2013.11.013 PMID:24480159

Tropsha, A. (2010). Best practices for QSAR model development, validation, and exploitation. *Molecular Informatics*, *29*(6-7), 476–488. doi:10.1002/minf.201000061

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Tropsha, A., Gramatica, P., & Gombar, V. K. (2003). The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR & Combinatorial Science*, *22*(1), 69–77. doi:10.1002/qsar.200390007

Vedani, A., Descloux, A. V., Spreafico, M., & Ernst, B. (2007). Predicting the toxic potential of drugs and chemicals in silico: A model for the peroxisome proliferator-activated receptor gamma (PPAR gamma). *Toxicology Letters*, *173*(1), 17–23. doi:10.1016/j.toxlet.2007.06.011 PMID:17643875

Vedani, A., & Dobler, M. (2002). 5D-QSAR: The key for simulating induced fit? *Journal of Medicinal Chemistry*, *45*(11), 2139–2149. doi:10.1021/jm011005p PMID:12014952

Vedani, A., Dobler, M., Dollinger, H., Hasselbach, K. M., Birke, F., & Lill, M. A. (2005). Novel ligands for the chemokine receptor-3 (CCR3): A receptor-modeling study based on 5D-QSAR. *Journal of Medicinal Chemistry*, *48*(5), 1515–1527. doi:10.1021/jm040827u PMID:15743194

Vedani, A., Dobler, M., & Lill, M. A. (2005). Combining protein modeling and 6D-QSAR. Simulating the binding of structurally diverse ligands to the estrogen receptor. *Journal of Medicinal Chemistry*, *48*(11), 3700–3703. doi:10.1021/jm050185q PMID:15916421

Vedani, A., Dobler, M., & Lill, M. A. (2006). The challenge of predicting drug toxicity in silico. *Basic & Clinical Pharmacology & Toxicology*, *99*(3), 195–208. doi:10.1111/j.1742-7843.2006.pto_471.x PMID:16930291

Vera-Divaio, M. A., Freitas, A. C., Castro, H. C., de Albuquerque, S., Cabral, L. M., & Rodrigues, C. R. et al. (2009). Synthesis, antichagasic in vitro evaluation, cytotoxicity assays, molecular modeling and SAR/QSAR studies of a 2-phenyl-3-(1-phenyl-1H-pyrazol-4-yl)-acrylic acid benzylidene-carbohydrazide series. *Bioorganic & Medicinal Chemistry*, *17*(1), 295–302. doi:10.1016/j.bmc.2008.10.085 PMID:19036592

Verloop, A. (1987). *The STERIMOL approach to drug design*. New York: Marcel Dekker.

Verma, J., Khedkar, V. M., & Coutinho, E. C. (2010). 3D-QSAR in drug design--A review. *Current Topics in Medicinal Chemistry*, *10*(1), 95–115. doi:10.2174/156802610790232260 PMID:19929826

von Itzstein, M., Wu, W. Y., Kok, G. B., Pegg, M. S., Dyason, J. C., & Jin, B. et al. (1993). Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature*, *363*(6428), 418–423. doi:10.1038/363418a0 PMID:8502295

Walters, D. E., & Hinds, R. M. (1994). Genetically evolved receptor models: A computational approach to construction of receptor models. *Journal of Medicinal Chemistry*, *37*(16), 2527–2536. doi:10.1021/jm00042a006 PMID:8057298

Wang, L., Liu, X., Shan, Z., & Shi, L. (2010). Using electrotopological state indices to model the depuration rates of polychlorinated biphenyls in mussels of Elliptio complanata. *Journal of Environmental Sciences (China)*, *22*(10), 1544–1550. doi:10.1016/S1001-0742(09)60287-4 PMID:21235184

Wang, R., Jiang, J., Pan, Y., Cao, H., & Cui, Y. (2009). Prediction of impact sensitivity of nitro energetic compounds by neural network based on electrotopological-state indices. *Journal of Hazardous Materials*, *166*(1), 155–186. doi:10.1016/j.jhazmat.2008.11.005 PMID:19101083

Warr, W. A. (2011). Representation of chemical structures. *Advanced Review*, *1*, 557–579.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Weininger, D. (1988). SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences*, *28*(1), 31–36. doi:10.1021/ci00057a005

Weininger, D., Weininger, A., & Weininger, J. L. (1989). SMILES. 2. Algorithm for generation of unique SMILES notation. *Journal of Chemical Information and Computer Sciences*, *29*(2), 97–101. doi:10.1021/ci00062a008

Wiener, H. (1947). Structural determination of paraffin boiling points. *Journal of the American Chemical Society*, *69*(1), 17–20. doi:10.1021/ja01193a005 PMID:20291038

Williams, M. S., Weiss, E. J., Sabatine, M. S., Simon, D. I., Bahou, W. F., & Becker, L. C. et al. (2010). Genetic regulation of platelet receptor expression and function: Application in clinical practice and drug development. *Arteriosclerosis, Thrombosis, and Vascular Biology*, *30*(12), 2372–2384. doi:10.1161/ATVBAHA.110.218131 PMID:21084706

Wold, H. (1966). *Research papers in statistics*. New York, NY: Wiley.

Wold, S., & Eriksson, L. (1995). Statistical validation of QSAR results. In H. Waterbeemd (Ed.), *Chemometric methods in molecular design* (Vol. 2, pp. 309–318). Weinheim: VCH. doi:10.1002/9783527615452.ch5

Xia, B., Liu, K., Gong, Z., Zheng, B., Zhang, X., & Fan, B. (2009). Rapid toxicity prediction of organic chemicals to Chlorella vulgaris using quantitative structure-activity relationships methods. *Ecotoxicology and Environmental Safety*, *72*(3), 787–794. doi:10.1016/j.ecoenv.2008.09.002 PMID:18950860

Young, D. C. (2009). *Computational drug design: A guide for computational and medicinal chemists*. John Wiely & Sons. doi:10.1002/9780470451854

Yuan, H., & Parrill, A. (2005). Cluster analysis and three-dimensional QSAR studies of HIV-1 integrase inhibitors. *Journal of Molecular Graphics & Modelling*, *23*(4), 317–328. doi:10.1016/j.jmgm.2004.10.003 PMID:15670952

Zarei, K., Atabati, M., & Kor, K. (2014). Bee algorithm and adaptive neuro-fuzzy inference system as tools for QSAR study toxicity of substituted benzenes to tetrahymena pyriformis. *Bulletin of Environmental Contamination and Toxicology*, *92*(6), 642–649. doi:10.1007/s00128-014-1253-2 PMID:24638918

Zhang, J., Rivard, B., & Rogge, D. M. (2008). The successive projection algorithm (SPA), an algorithm with a spatial constraint for the automatic search of endmembers in hyperspectral data. *Sensors (Basel, Switzerland)*, *8*(2), 1321–1342. doi:10.3390/s8021321

Zhang, S., Wei, L., Bastow, K., Zheng, W., Brossi, A., Lee, K. H., & Tropsha, A. (2007). Antitumor agents 252. Application of validated QSAR models to database mining: Discovery of novel tylophorine derivatives as potential anticancer agents. *Journal of Computer-Aided Molecular Design*, *21*(1-3), 97–112. doi:10.1007/s10822-007-9102-6 PMID:17340042

Zhu, H. (2013). From QSAR to QSIIR: Searching for enhanced computational toxicology models. *Methods in Molecular Biology (Clifton, N.J.)*, *930*, 53–65. doi:10.1007/978-1-62703-059-5_3 PMID:23086837

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

KEY TERMS AND DEFINITIONS

3D-QSAR: Quantitative analysis of relationship between spatial features of a set of compounds and biological activities using statistical methods.

Accuracy: The accuracy of a predicted model, the most important quality measurement, is usually calculated for both training and test sets.

Dimensionality Reduction: A transformation function/algorithm which maps a large dataset of descriptors in to a small dataset of descriptors as a representative subset of the large dataset. The algorithms can be either linear or non-linear.

Machine Learning: An important field of artificial intelligence in computer science is the machine learning. Machine learning algorithms are diverse and are used to extract the rules of black box functionality by mapping the inputs to the corresponding outputs using the computational and statistical calculations.

Molecular Descriptor: Numerical representation of molecular properties (e.g. physicochemical properties) of chemical structures.

Pharmacophore: Minimum structural requirement of a compound necessary for interaction with a target molecule.

QSAR: Quantitative correlation between biological activities of compounds and chemical structures represented by a regression model.

Test Data Set: A set of compounds which does not have any role in training process and is used for model evaluation.

Training Data Set: A set of molecules selected for training of the model using machine learning algorithms.

ENDNOTES

1 Tripos is a well-known and global pioneer in computational drug discovery capabilities by designing and development of scientific software products. http://www.tripos.com

2 Advanced Chemistry Development, Inc., (ACD/Labs) is one of the companies providing chemistry software for data management and informatics. ACD/Labs suite of programs provide *in silico* prediction of molecular physical properties and ADMET property. ACD/Labs Release 2014 (v2014) software is the latest version.http://www.acdlabs.com/

3 CambridgeSoft was a cheminformatics software company. In 2011, the company was sold to PerkinElmer, a laboratory instruments company. The collection of programs supplies the chemical structures and their associated data and properties.http://www.cambridgesoft.com/

4 InfoChem is a collection of software tools and applications for chemical data management, storage, handling and retrieval of structures and reactions.http://www.infochem.de/

5 Dragon was developed by an Italian company, namely, Talete. It is designed for calculation of molecular descriptors which can be used in QSAR studies. The latest version is Dragon 6. It can be found at:http://www.talete.mi.it/products/dragon_description.htm

6 PubChem was released in 2004 as a database within national center for biotechnology information (NCBI) for providing information on chemical structure and biological activities of small molecules.

7 ChemDraw is a chemical drawing tool for generating images of manuscripts, databases, and making useful queries in chemical databases

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

https://www.cambridgesoft.com/Ensemble_for_Chemistry/ChemDraw/

[8] ACD/ChemSketch is a drawing package that includes drawing and calculation of chemical structures and molecular properties, and prediction of logP.
http://www.acdlabs.com/resources/freeware/chemsketch/

[9] Open Babel is a free chemical toolbox designed to search, convert, analyze, or store data from molecular modeling and related areas. http://openbabel.org/wiki/Main_Page

[10] CORINA is useful in generating 3D structure for small and medium sized molecules.
https://www.molecular-networks.com/products/corina

[11] Concord which is included in SYBLE is used for equality assessments on every structures in 3D.

[12] Frog (FRee Online druG) is a tool for generating small molecular conformations by using 1D or 2D descriptions. http://bioserv.rpbs.jussieu.fr/cgi-bin/Frog

[13] Smi23d or (3D Cordinate Generation) program is able to convert one or more SMILIES into 3D structure.
http://www.chembiogrid.org/cheminfo/smi23d/

[14] ADRIANA.Code is used to compute molecular structure descriptors in drug design and predict the ADME/toxic properties. https://www.molecular-networks.com/products/adrianacode

[15] Molconn-Z is a software which creates Molecular Connectivity, Shape, and Information Indices for (QSAR) Analyses. The new E-State parameter is also included. http://www.edusoft-lc.com/molconn/

[16] A software that calculates 1875 molecular descriptors (i.e., 1444 1D, 2D descriptors and 431 3D descriptors and fingerprints). http://padel.nus.edu.sg/software/padeldescriptor/

[17] KNIME is an open source software for intelligent algorithms including the data mining, machine learning, and others. https://www.knime.org/

[18] RapidMiner is mostly used for data integration, data transformation, data modeling, and data visualization.
http://rapidminer.com/

[19] Weka (Waikato Environment for Knowledge Analysis) is a general machine learning software including some common statistical, clustering, and intelligent system algorithms written in Java.
http://www.cs.waikato.ac.nz/ml/weka/

[20] An Open source program based on visual programming or Python used for data visualization and analysis.
http://orange.biolab.si/

[21] TANAGRA is a free software which includes some data mining methods and is capable of analyzing the through statistical and, machine learning algorithms. http://eric.univ-lyon2.fr/~ricco/tanagra/en/tanagra.html

[22] MATLAB (matrix laboratory) is a numerical computing environment which inclides various types of toolboxes for artificial neural networks, image processing, bioinformatics, parallel computing, optimization procedures and etc. http://www.mathworks.com/products/matlab/

[23] R is a free software environment alike MATLAB for statistical and machine learning computing and graphical illustrations. Many packages over the world are being written for the purpose of general and expert researches. http://www.r-project.org/

[24] SYBYL® is considered as the heart of Tripos. This software has various applications including construction, editing, and visualization tools, modeling and simulation of the molecules and lead identification, and lead optimization. SYBYL-X is the latest version.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

[25] Discovery Studio is developed by Accelrys used on small/macro molecule and simulation. http://accelrys.com/products/discovery-studio/

[26] MOE is a software used generally for life and material sciences by comprehensively performing the visualization, simulations on molecular structures. http://www.chemcomp.com/MOE-Molecular_Operating_Environment.htm

[27] CODESSA is a tool for obtaining descriptors to develop QSAR/QSPR model through quantum mechanical results from sources such as AMPAC, Gaussian, AIMALL. http://www.semichem.com/codessa/

[28] It is a 3D-QSAR program calculating of total interaction energies between different chemical probes and molecules. It determines energetically favorable binding sites on known molecular structures as well as surface properties. GRID 22c is the latest version. http://www.moldiscovery.com/software/grid

[29] Molecular Discovery is one of the companies in the field of structure-based drug design. All the developed programs, based on GRID force field, are designed for 3D-QSAR studies, virtual screening, scaffold-hopping, ADME and pharmacokinetic modelling, optimization of metabolic stability and metabolite prediction. http://www.moldiscovery.com/

[30] Pentacle is a alignment-independent 3D-QSAR software. It computes interaction energies between the molecule and chemical probes (GRID based Molecular Interaction Fields, or MIFs). The latest version of Pentacle is 1.0.7.http://www.moldiscovery.com/software/pentacle

[31] GOLPE (Generating Optimal Linear PLS Estimations) is an advanced chemometric toolbox for building, validating and interpreting the 3D QSAR models. http://www.miasrl.com/golpe.htm

[32] The toolbox is especially suitable for chemical researchers to assess the toxicity of the chemical compounds. The latest version is 3.2. http://www.oecd.org/chemicalsafety/risk-assessment/theo-ecdqsartoolbox.htm

[33] REACH stands for Registration, Evaluation, Authorisation and Restriction of Chemicals which has been approved by European Union in 2006 for assessing the chemical effects on health and environment.

[34] TOXNET (TOXicology Data NETwork) is a group of toxicology dataset in various sections including the chemicals, drugs and environmental effects. http://toxnet.nlm.nih.gov/

[35] SIDER stands for Side Effect Resource which includes the information about side effect frequency of drugs. And the latest version of SIDER is 2.0 and can be found at the link ttp://sideeffects.embl.de/

[36] ACToR is a tool for the interested researches in chemical toxicity to extract data about the possible risk of current chemicals on human health. http://www.epa.gov/actor/

[37] DailyMed provides the most comprehensive information about the marketed drugs which are approved by Food and Drug Administration (FDA). http://dailymed.nlm.nih.gov/dailymed/about.cfm

[38] Distributed Structure-Searchable Toxicity (DSSTox) Database Network is a project of EPA for publicly presenting the data about structure-activity and predictive toxicology capabilities http://www.epa.gov/ncct/dsstox/.

This work was previously published in Quantitative Structure-Activity Relationships in Drug Design, Predictive Toxicology, and Risk Assessment edited by Kunal Roy, pages 1-47, copyright year 2015 by Medical Information Science Reference (an imprint of IGI Global).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 3
The History and Development of Quantitative Structure–Activity Relationships (QSARs)

John C. Dearden
Liverpool John Moores University, UK

ABSTRACT

It is widely accepted that modern QSAR began in the early 1960s. However, as long ago as 1816 scientists were making predictions about physical and chemical properties. The first investigations into the correlation of biological activities with physicochemical properties such as molecular weight and aqueous solubility began in 1841, almost 60 years before the important work of Overton and Meyer linking aquatic toxicity to lipid-water partitioning. Throughout the 20th century QSAR progressed, though there were many lean years. In 1962 came the seminal work of Corwin Hansch and co-workers, which stimulated a huge interest in the prediction of biological activities. Initially that interest lay largely within medicinal chemistry and drug design, but in the 1970s and 1980s, with increasing ecotoxicological concerns, QSAR modelling of environmental toxicities began to grow, especially once regulatory authorities became involved. Since then QSAR has continued to expand, with over 1400 publications annually from 2011 onwards.

INTRODUCTION: WHAT IS A QSAR?

Humans are inherently inquisitive. Even small children persistently ask "Why?". So it is no surprise that for many years scientists have asked why some substances have a beneficial effect on the body, whilst others are toxic, and why some are more beneficial, or more toxic, than are others. That led Crum Brown and Fraser (1868-1869) to postulate that "there can be no reasonable doubt but that a relation exists between the physiologic action of a substance (Φ) and its chemical composition and constitution (C)". Hence $\Phi = f$C. They did not go on to suggest what functions of composition and constitution might be important. Nevertheless, their equation is a valid generic quantitative structure-activity relationship (QSAR). They also pointed out that "to discover *f* we produce a known change on the constitution by

DOI: 10.4018/978-1-5225-0549-5.ch003

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

which it becomes $C + \Delta C$, and examine the corresponding change of physiological action which has become $\Phi + \Delta\Phi$. We thus obtain the relation between ΔC and $\Delta\Phi$, and by sufficiently varying C and ΔC, we may hope to get at all events an approximate solution of the problem". That was a remarkably prescient statement, for it is exactly how QSAR modelling is performed (Kubinyi, 1993).

Hence, in Crum Brown and Fraser's terminology, a QSAR equation would be:

$$\Phi = c_1C_1 + c_2C_2 + c_3C_3 + \ldots c_nC_n \quad (1)$$

where C_n represents one constitutional (structural) property, and c_n is its coefficient.

A QSAR is now defined as a mathematical relationship linking chemical structure and pharmacological activity or other property in a quantitative manner for a series of compounds. It should be noted that a correlation between a physico-chemical property (such as aqueous solubility) and some function(s) of chemical composition and constitution is usually called a quantitative structure-property relationship (QSPR), although strictly the term QSPR covers both structure-biological activity and structure-physico-chemical property relationships, as the title of this journal indicates.

As will be seen, the growth of QSAR since the 1970s has been huge. Consequently, it has not been possible in this brief review to provide more than a glimpse into modern aspects and achievements of QSAR. For those wishing to know more of those approaches, the recent excellent review by Cherkasov et al. (2014) is recommended.

EARLY APPROACHES

It is significant that for centuries there has been recognition that quantitation is an essential part of science. Leonardo da Vinci (1452-1519) commented that "there is no certainty in sciences where one of the mathematical sciences cannot be applied" (da Vinci). According to Galileo (1564-1642) "to study a given phenomenon, it was necessary to measure quantities, identify regularities, and obtain relationships representing mathematical descriptions as simply as possible" (Ponte, 1992). Gay-Lussac (1778-1850) optimistically said that "we are perhaps not far from the epoch when we will be able to submit to calculation the majority of chemical phenomena" (Gay-Lussac, 1809). Charles Babbage (1791-1871), the father of computing, could almost be thought to have had QSAR in mind when he stated that "all of chemistry…would become a branch of mathematical analysis which, like astronomy, taking its constants from observation, would enable us to predict the character of any new compound" (Babbage, 1837). A dissenting voice was that of Auguste Comte (1798-1857), who wrote that "every attempt to employ mathematical methods in the study of chemical questions must be considered profoundly irrational and contrary to the spirit of chemistry. If mathematical analysis should ever hold a prominent place in chemistry – an aberration which is happily almost impossible – it would occasion a rapid and widespread degeneration of that science" (Liang, Kvalheim, & Manne, 1993).

Sir William Thomson (1824-1907), later Lord Kelvin, was very forthright in his views on the importance of mathematics in science; "I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the state of *Science*, whatever the matter may be" (Thomson, 1884).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

THE BEGINNINGS OF CORRELATION (1816-1900)

Physicochemical Properties: The Periodic Table

So far as is known, the earliest work on property prediction involved the elements, or as many of them as were known in the early nineteenth century. The first to publish on this topic appears to have been Döbereiner, who in a letter to Goethe in 1816 (Kauffman, 1999) mentioned what was to evolve into his *Dreiheit* (rule of triads); "The mineral coelestine [strontium sulphate] shows remarkable relationships: its specific weight is the mean of that of [calcium sulphate] and [barium sulphate], namely (2.95 + 4.47)/2 = 3.71". The rule of triads is in effect a read-across technique (Van Leeuwen, Schultz, Henry, Diderich, & Veith, 2009) whereby a property value of a chemical is predicted from known values of that property from one or more similar chemicals. Nonetheless, it is a valid predictive approach, and thus can be included as an historical use of QSPR.

Some of Döbereiner's publications on the subject (Döbereiner, 1817, 1829) extended his ideas to many other families of elements, and in fact his was the first of numerous attempts to produce a systematic arrangement of the elements. A good and concise summary of those attempts has been given by van Spronsen (1959). The elucidation of patterns of elements was not helped by confusion over exact (or even approximate) values of atomic weights. Cooke (1854) developed a classification based on simple formulae for six different series, several of which were based on the number 8, rather than on triads. Odling (1857) and Dumas (1857) were able to make connections between what were formerly considered independent groups of elements. Zapffe (1969) stated that "it could easily be argued that Dumas was the initiator of the whole Mendeleevian stream of thought. He even went so far as to anticipate Mendeleev on the matter of predicting undiscovered elements where gaps appeared in the relationships". Newlands (1864) propounded his "law of octaves", observing an eightfold regularity among the elements. However, the first person actually to develop a periodic system of the elements was Béguyer de Chancourtois (1862), who formulated a helical pattern of groups of elements. Another significant step was taken by Hinrichs (1867) with his hand-written *Atomechanik* giving a group classification of the elements. In 1869 Hinrichs published details of his classification (Hinrichs, 1869).

It is surprising that, with all that went on in connection with classification of the elements prior to that of Mendeleev, his is the chief, and sometimes the only, name that is recalled. That is not to belittle his achievements, which were outstanding. He formulated the Periodic Law, created his own version of the periodic table of elements, and used it to correct the properties of some already discovered elements and also to predict the properties of eight elements yet to be discovered (Mendeleev, 1869). He is reported (Wikipedia) as saying: "I saw in a dream a table where all elements fell into place as required. Awakening, I immediately wrote it down on a piece of paper, only in one place did a correction later seem necessary."

Other Physicochemical Properties

Kopp (1842) observed regularities in the specific volumes, densities and boiling points of compounds within homologous series. His numbering is hard to follow, because he did not define units, and his atomic weights, whilst no doubt in line with the knowledge of his time, were wildly erroneous; for example, he gave the atomic weight of carbon as 75.854, and he appears to use "atomic weight" also to mean "molecular weight". Concerning boiling points, he reported for example that those of the oxalates are

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

all 106° higher than those of the acetates, and that substitution in any series of a methyl group by ethyl raised the boiling point by 18°. Since he looked only at relatively low molecular weight compounds, he failed to realise that within homologous series the boiling point difference between adjacent homologues (Δbp) decreases as chain length increases.

In a later publication Kopp (1844) gave more details of his calculation of specific volumes. For example, he obtained the specific volumes (units unspecified) at 35ºC of carbon (72.10), hydrogen (27.04) and oxygen (54.08), and was then able to calculate the specific volume of diethyl ether as 612.9, which he claimed agreed exactly with the measured value. This early approach is similar to that of McGowan (1986) for the calculation of "characteristic volumes".

Mills (1884a) actually formulated a quantitative structure-property relationship to model the melting and boiling points of a number of homologous series, which took account of the decrease in Δbp with increasing chain length. His generic QSPR was:

$$\text{mp or bp} = \frac{\beta\left(x-c\right)}{1+\gamma\left(x-c\right)} \tag{2}$$

where x = number of CH_2 groups in the chain, and β, γ and c are constants. Mills devised separate equations for odd- and even-numbered chains, and found that for many series melting and boiling points could be predicted to within 1 or 2 degrees. This can probably be considered to be the first proper QSPR.

Mills (1884b) also attempted to model the periodicity of the elements, using a rather complicated formula. He did not appear to be aware of the work of Mendeleev (1869)!

Biological Activities

Blake (1841) reported the toxic effects of inorganic salts on animals such as dogs and rabbits; he found that within what he termed isomorphous groups of metal salts (which in general resemble the divisions later proposed for the Periodic Table (*vide ultra*)) "the action of substances in the same isomorphous group is a function of the atomic weight (of the metal), the greater the atomic weight the smaller the quantity required to produce the same physiological action". In a later work (Blake, 1884) he expanded his studies to include forty elements, but commented that "with these extended investigations no facts have been met with that were not in accord with the conclusions I then (i.e. in 1841) published".

In a M.D. thesis at the University of Strasbourg, Cros (1863) reported an increase in the toxicity of alcohols with decreasing aqueous solubility, up to a maximal potency, limited presumably by very low solubility of the higher alcohols. Similar variations of toxicity with aqueous solubility were observed by Rabuteau (1870), Dujardin-Beaumetz & Audigé (1876) and Richet (1893), but none cited earlier works (Lipnick, 1990). Richardson (1869) observed that the narcotic effect of primary alcohols in several species increased in proportion to their molecular weight. Brunton (1880) expressed the hope that physicians would soon be able to predict the pharmacological action of any compound from its chemical constitution. It is a sombre thought that even now, 136 years on, that goal has not yet been fully achieved. As pointed out above, Crum Brown and Fraser (1868-1869) showed that the pragmatic approach was to examine how changes in chemical constitution (structure) could produce changes in biological response. Bynum (1970) has reviewed much of the 19th century work on the relationship between chemical structure and pharmacological action, and Parascandola (1971) has given more details of the work of Blake, Rich-

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

ardson, Crum Brown & Fraser, and Brunton. It is pleasing to note that Richardson, Brunton and Fraser were all knighted for their services to chemistry. Walker, Enache and Dearden (2003) have reviewed the early work on structure-toxicity relationships of metals and metal ions.

What those who related toxicity to aqueous solubility and molecular weight failed to recognise is that both properties were surrogates for the partition coefficient. Hansch et al. (1968) showed that aqueous solubility of liquids was inversely related to partition coefficient. For homologous series, with successive members differing by CH_2, molecular weight is directly proportional to the logarithm of the partition coefficient. The significance of partition coefficient cannot be over-emphasised; it is undoubtedly the main factor controlling the transport of a chemical within an organism, as it repeatedly crosses lipid membranes on its "random walk" to a site of action (Hansch, 1969).

The breakthrough – not too strong a word – came with the work of Overton (1897, 1899) and Meyer (1899), working independently. They found that the toxicity of narcotic compounds to tadpoles and other aquatic species could be correlated with their oil-water partition coefficients; olive oil was the main oil that they used. Overton's book, *Studien über die Narkose* (1901) has been translated, with additions, into English by Lipnick (1991).

Although neither Overton nor Meyer produced a correlation equation, Hansch (1969) used some of Overton's tadpole data in a QSAR (equation 3), using the 1-octanol-water partition coefficient (P), whilst Leo et al. (1969) found that the Overton data correlated only poorly with other physicochemical properties (polarisability, molar attraction, parachor and molecular weight).

$$\log 1/C = 0.94 \log P + 0.87 \quad (3)$$

$$n = 51\ r^2 = 0.943.$$

where C = concentration to produce narcosis in a given time, and n = number of compounds used to develop the model.

The high coefficient of variation (r^2) is tribute to the accurate work of Overton.

MORE PROGRESS (1901-1962)

The level of QSAR/QSPR type activity in the late 19th century continued into the early 20th century. In 1903 Longinescu derived the following expression for the melting points of organic compounds (Godavarthy, Robinson, & Gasem, 2006):

$$T_m = 8.37D\sqrt{n} \quad (4)$$

where D = density, and n = number of atoms in molecule.

A number of other models for melting point prediction appeared over the years, and these have been discussed by Dearden (1999).

Führer (1903, 1904) found that the toxicity of alcohols to sea-urchin eggs increased with molecular weight, up to a limit imposed by lack of aqueous solubility. He also pointed out that within homologous series narcotic toxicities increased in a geometric progression: 1, 3, 3^2, 3^3 and so on, suggesting additivity of group contributions.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Using some of Overton's data, Traube (1904) found that the toxicity of 11 simple organic compounds increased with the surface tension of the compounds in aqueous solution, and proposed that this related to the ability to penetrate membranes. Vernon (1911) found that the ability of a series of 10 alcohols to lower tortoise heart rate ranked well with Overton's data for tadpole narcosis by the same alcohols.

Holt (1916) found that the toxicity of volatile organic compounds to the cockroach increased as boiling point increased up to a certain point, beyond which toxicity fell again. A similar result was observed by Moore (1917a), who also examined the relationship between toxicity to insects and volatility (Moore, 1917b) and, for liquid insecticides, their surface tension (Moore, 1918). Warburg (1921) investigated narcosis of avian blood cells by simple organic compounds. He considered that narcotics acted by surface properties, and, like Traube before him (Traube, 1904), found that narcosis appeared to be related to surface tension.

Numerous publications in the 1920s and 1930s sought to relate biological activity to molecular weight or aqueous solubility. They include the work of Tilley & Schaffer (1926, 1928), Dohme et al. (1926), and Klarmann et al. (1931). The latter two found activity to pass through a maximum as molecular weight increased. But it is remarkable that, despite the work of Overton and Meyer at the end of the 19th century (*vide ultra*), there appears to be not a single publication in the 1920s and 1930s that used oil-water partition coefficient as a key descriptor in modelling biological activities.

Much of the work cited above was quantified by Ferguson (1939), who showed that the toxic concentrations (C_T) of a series of compounds could be predicted from their aqueous solubility or vapour pressure using equation 5:

$$C_T = kA^{1/n} \tag{5}$$

where k and n are constants, with n >1, and A = solubility or vapour pressure. Later he developed a generalised form of equation 5 in which k can be another physicochemical property such as partition coefficient (Ferguson & Pirie, 1948). Purcell et al. (1973) have pointed out that equation 5 can be transformed logarithmically into the format now widely used in QSAR/QSPR work:

$$\log 1/C = k' + n'.\log A \tag{6}$$

Ferguson (1939) also pointed out that "physiological effects due to a physical mechanism are generally measured when an equilibrium has been established between the concentration of the toxic substance in the phase, solution or vapour, in which it is applied and the concentration in the phase or surface layer which is the seat of toxic action. Since the chemical potential of the toxic substance must be the same in all phases partaking in the equilibrium, it is suggested that the toxicities of physically toxic substances should be compared, not by the value of the toxic concentration in the external solution or vapour, but by the values of the chemical potentials in these phases. Chemical potentials so determined are identical with the chemical potentials at the actual point of attack within the organism. Using the activity function of G.N. Lewis as the chemical potential, it is found that, when chemical reaction is absent, though diverse chemical compounds exert the same toxic effect on a given organism at widely different concentration, the activities corresponding to these concentrations lie within a relatively narrow range". For example, for a series of 14 narcotic gases and vapours tested on mice, whilst the narcotic concentrations ranged from 0.5% - 100%, the corresponding activities ranged only from 0.01 to 0.07.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Another great step forward was the development of electronic substituent constants by Hammett (1935, 1937), which could be correlated with reaction and equilibrium constants within a chemical series using equation 7:

$$\log (k_X/k_H) = \rho\sigma \quad (7)$$

where k_H is the reaction rate or equilibrium constant of the parent compound and k_X is that of a derivative with substituent X; ρ is the series constant, and σ is what is now called the Hammett substituent constant. Hammett's original work involved the acid dissociation constants of a series of benzoic acids, for which he set ρ equal to unity. For other series and other equilibria or rate constants, ρ has different values. Hammett's work set in train a new field of study called linear free energy relationships (LFER). This name derives from the van't Hoff isotherm, which shows that the free energy change in a process is directly proportional to the logarithm of the equilibrium or rate constant of the process:

$$\Delta G = -2.303RT\log K \quad (8)$$

where ΔG = Gibbs free energy, R = universal gas constant, T = absolute temperature, and K = an equilibrium or rate constant. Following Hammett, much work was done in modelling physical and chemical processes using the Hammett substituent constant (Wells, 1968). However, it was soon recognised that σ has two components, resonance (R) and inductive (I) effects; furthermore, it is position-dependent. Hansch and Leo (1979) have discussed these and other effects in some detail, and give an example of the modelling of pKa values of benzoic acids:

$$pKa = 5.76 - 1.64\,\sigma_I^m - 0.34\,\sigma_R^m - 1.76\,\sigma_I^p - 1.15\,\sigma_R^p \quad (9)$$

$$n = 27\ r^2 = 0.984\ s = 0.083$$

where superscripts m and p refer to meta- and para-substitution. Hansch and Leo (1979) also included in their book many values of a number of types of Hammett constants.

Hammett constants model electronic effects, and hence Taft (1952) devised a steric constant, E_s, in order to incorporate steric effects into LFERs.

Although the Hammett substituent constant was developed and used successfully for correlating physicochemical properties, it is not surprising that attempts were made to examine whether it could be used in the correlation of biological activities. For example, Hansen (1962) reported the toxicities and σ constants of benzoic acids to mosquito larvae, from which Kubinyi (1993) developed the following QSAR:

$$\log 1/C = 1.787 + 1.454\,\sigma \quad (10)$$

$$n = 13\ r^2 = 0.843\ s = 0.243$$

Zahradnik and Chvapil (1960) also attempted to correlate various biological activities with Hammett and Taft constants, with only limited success. Jaffé (1953) commented that he had not found any series of compounds whose biological activity could be described by Hammett constants.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

A number of boiling point correlations published in the first half of the 20th century have been listed by Egloff et al. (1940); they also published their own QSPR for the prediction of boiling points (T) of normal alkanes:

$$T = 745.42 \log (n + 4.4) - 416.31 \quad (11)$$

where n = number of carbon atoms. Other than for methane, the prediction errors averaged 0.40°. Surprisingly, Egloff et al. (1940) did not mention the excellent correlations of Mills (1884a).

A very important contribution, almost totally ignored, was that of Bell and Roblin (1942), who investigated the bacteriostatic activities of a series of sulfanilamides. They found a biphasic relationship between activity and pKa, and developed a QSAR to model that relationship:

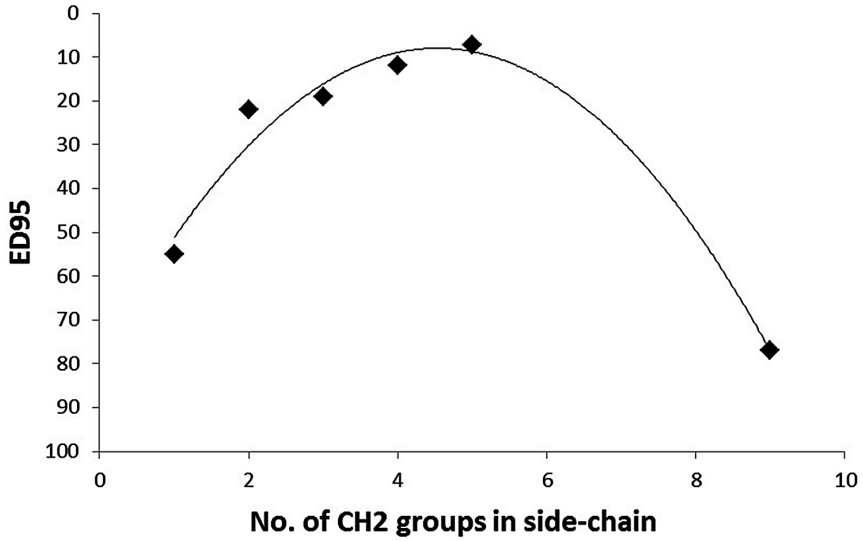
$$\log (0.001/C_R) = 3.23 + 0.661 \log x + 0.339 \log (1 - x) \quad (12)$$

where C_R = minimum molar concentration to exhibit a given level of bacteriostatic activity, x = fraction of the total concentration in ionic form.

Several things are especially noteworthy about this work. Firstly, a good biphasic correlation was reported. Secondly, equation 12 is very similar to the modern format of QSAR equations, including the use of logarithmic functions, the use of reciprocal concentrations and the use of molar concentrations. Thirdly, the title of the paper includes the phrase "relation of structure to activity", which may be the origin of the now widely used term "structure-activity relationship". Equation 12 is quite probably the very first biological QSAR.

Fieser and Richardson (1948) found that the activities of several series of naphthoquinone antimalarials in ducks varied parabolically with the number of carbon atoms in the side-chain (see Figure 1). Unfortunately they gave no mechanistic interpretation of their observations, although in a later publication

Figure 1. Anti-malarial activity of cyclohexylalkylnaphthoquinones against P. lophurae in ducks (after Fieser & Richardson, 1948)



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

(Fieser, Ettlinger, & Fawaz, 1948) they reported that the values of log P of the most active members of eight series of naphthoquinones were all in the range 10 – 12.

A quadratic equation fitting the data shown in Figure 1 has good statistics ($r^2 = 0.974$); when the data are transformed to the logarithm of molar concentrations, as is the correct QSAR practice, the correlation is still good ($r^2 = 0.952$).

McGowan (1952) devised a QSAR for the toxicities of diverse narcotic compounds that gave good correlations:

$$\log C = S + (4 \times 10^{23})\, p - 0.036\, P_c + 7.7 \qquad (13)$$

where S = sensitivity factor based on varying sensitivity of different biological systems, p = polarisability, and P_c = parachor. It should be noted that both polarisability and parachor have units of molecular volume, so in effect McGowan was correlating narcotic toxicity with molecular size, as had many others before him.

Rather than trying to model biological activities using Hammett constants, Bruice et al. (1956) developed empirical substituent constants based on the contribution of each substituent to thyromimetic activity in both amphibians and mammals, and obtained good correlations. The problems with this type of analysis are that new constants have to be derived for each type of biological activity, that the contribution of each substituent is fixed, and that mechanistic interpretation is difficult.

1962 – PRESENT

It was not until 1962 – 65 years after Overton (1897) first showed that oil-water partitioning was the key process in narcosis – that the second major breakthrough occurred. Corwin Hansch, of Pomona College, California, and co-workers (Hansch, Maloney, Fujita, & Muir, 1962) showed that, for a series of phenoxyacetic acid plant growth regulators, the activity in *Avena* (oats) could be correlated with both the Hammett substituent constant σ and a hydrophobic substituent constant π:

$$\log 1/C = 3.36 + 4.08\, \pi - 2.14\, \pi^2 + 2.78\, \sigma \qquad (14)$$

where C = concentration to produce a specified growth in *Avena*.

This seminal work was outstanding in several ways:

1. It showed that complex biological effects, and not simply narcosis, could be modelled by physicochemical descriptors;
2. It showed that more than one type of descriptor was sometimes required to obtain a good model of biological activity;
3. It showed that biological activity can be modelled by a biphasic correlation. As pointed out above, others (Dohme, Cox & Miller, 1926; Klarmann, Gatyas & Shternov, 1931; Bell & Roblin, 1942; Fieser & Richardson, 1948) had found that a plot of biological activity against a physicochemical descriptor often rose to a maximum and then fell again, but no-one had previously represented that by a quadratic equation;

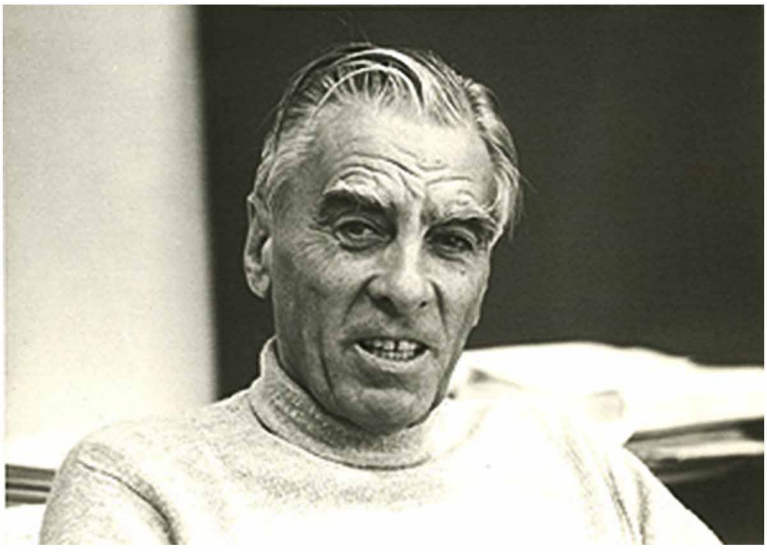
EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

4. It pioneered the use of 1-octanol as a surrogate for lipid. Whilst the analogy is not perfect, both 1-octanol and lipids typically have a polar end-group and a hydrocarbon chain, and octanol-water π and log P values have been found to be valuable descriptors in very many QSAR correlations. Other pinguid phases were later proposed (Leahy, Morris, Taylor, & Wait, 1992), but 1-octanol still remains the most widely used;
5. It pioneered the use of the hydrophobic substituent constant, π, which is defined for a substituent X on a molecule MX as $\pi_X = \log(P_{MX}/P_{MH})$ or $\log P_{MX} - \log P_{MH}$, where P is the octanol-water partition coefficient. For example, the log P of toluene is 2.68, and that of benzene is 2.13, so π_{CH3} is 2.68 – 21.3 = 0.55.This in effect means that hydrophobicity is, to a first approximation at least, an additive property.

However, it is not quite correct to say that benzene + methyl = toluene; it is phenyl (C_6H_5) + methyl that equals toluene. Hence the calculation of log P of toluene as $\log P_{benzene} + \pi_{CH3}$ is wrong by π_H. This is a negligible difference, but for a long-chain compound the error becomes significant. This was pointed out by Nys and Rekker (1973), who introduced a different method of calculating log P values. By analysing the log P values of a large number of compounds, they were able to calculate the contribution of molecular fragments (hydrophobic fragmental constants) to log P. This method avoided the problem found with the hydrophobic substituent constant, as was acknowledged by Leo et al. (1975), who introduced their own fragmental constants, obtained from the careful measurement of log P values of very small molecules such as hydrogen and methane. This work formed the basis for the ClogP software for the calculation of log P.

Hansch and Fujita (1964) expanded their studies by showing that their approach worked with, for example, the toxicity of phenols to bacteria, the toxicity of benzoic acids to mosquito larvae, and the toxicity of diethyl phenyl phosphates to houseflies. Since then the Hansch group have published many

Figure 2. Professor Corwin Hansch, father of QSAR: 6 October 1918 – 8 May 2011 (photograph courtesy of the Department of Chemistry, Pomona College, Pomona, CA)



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

papers covering a wide range of chemicals and species, and Hansch (1969) has presented a succinct account of what has come to be called Hansch analysis (Kubinyi, 1993). Hansch and Dunn (1972) and Hansch and Clayton (1973) reported numerous examples of biological activities correlated with log P using rectilinear and parabolic models respectively. Others (van de Waterbeemd, 1992; Kubinyi, 2002; Selassie, 2003; Martin, 2012) have reviewed the QSAR contributions of Hansch and his co-workers. It is not for nothing that Corwin Hansch has been called the father of QSAR (Martin & Stouch, 2011).

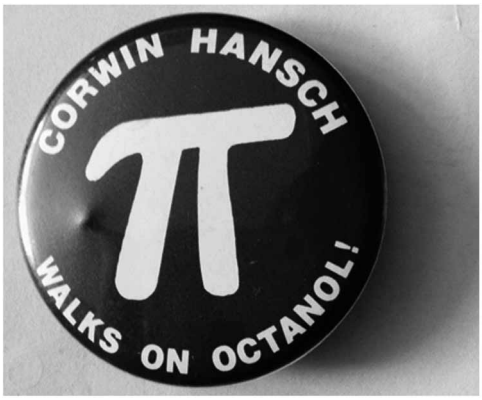
Contemporaneously with the ρ-σ-π approach of Hansch and Fujita (1964), Free and Wilson (1964) developed an approach similar to that of Bruice et al (1956), using indicator variables for the presence of structural features different from those of a chosen reference compound. These were correlated with the relevant biological activity data, and the regression coefficients generated represent the contributions of each structural feature to the model. Kubinyi (1993) has commented that the Free-Wilson approach is difficult to use, and the Fujita and Ban modification (1971) offers important advantages. However, the approach is little used today. Kubinyi (1993) has given a detailed analysis of the Free-Wilson and Fujita-Ban approaches.

Most molecules are flexible, and hence are able to adopt a number of conformations. When a chemical binds to a receptor site in an organism, it does so in what can be considered a preferred conformation, which is usually a low-energy conformation. Kier (1967) utilized extended Hückel theory molecular orbital calculations to obtain preferred conformations, and found good agreement with conformations obtained from crystal X-ray analysis. He later reported many other such calculations, for example for thyrotropin-releasing hormone (George & Kier, 1973). Since then many other workers have utilized M.O. theory to generate preferred conformations for QSAR modeling (Martin, 2002).

In the 1970s and 1980s, a story was circulating in QSAR circles that a car had been seen on a U.S. highway with a bumper sticker saying: “Corwin Hansch walks on octanol”. At the 1988 European QSAR Symposium, held in Interlaken, Switzerland, when Hansch was 70 years old, lapel badges were given to all delegates with those words on (see Figure 3).

Many scientists found it hard to accept that something as complex as, for example, anti-inflammatory activity of drugs, or insecticidal activity, can be modelled by a few physicochemical and/or structural properties of chemicals. In fact, what QSAR does is to model the change of biological activity with change of chemical structure, as Crum Brown and Fraser (1868-1869) pointed out many years ago. Most biological processes are multi-step, but generally one of those steps will be rate-limiting, and so the whole

Figure 3. Lapel badge given to delegates attending the 1988 European QSAR Symposium



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

process can be modelled with (often simple) chemical properties. An example is the anti-inflammatory activity in the rat of a series of aspirin derivatives (Dearden & George, 1979), which is modelled by the 1-octanol-aqueous buffer (pH1) partition coefficient alone, as Figure 4 shows.

The QSAR for the anti-inflammatory activity is:

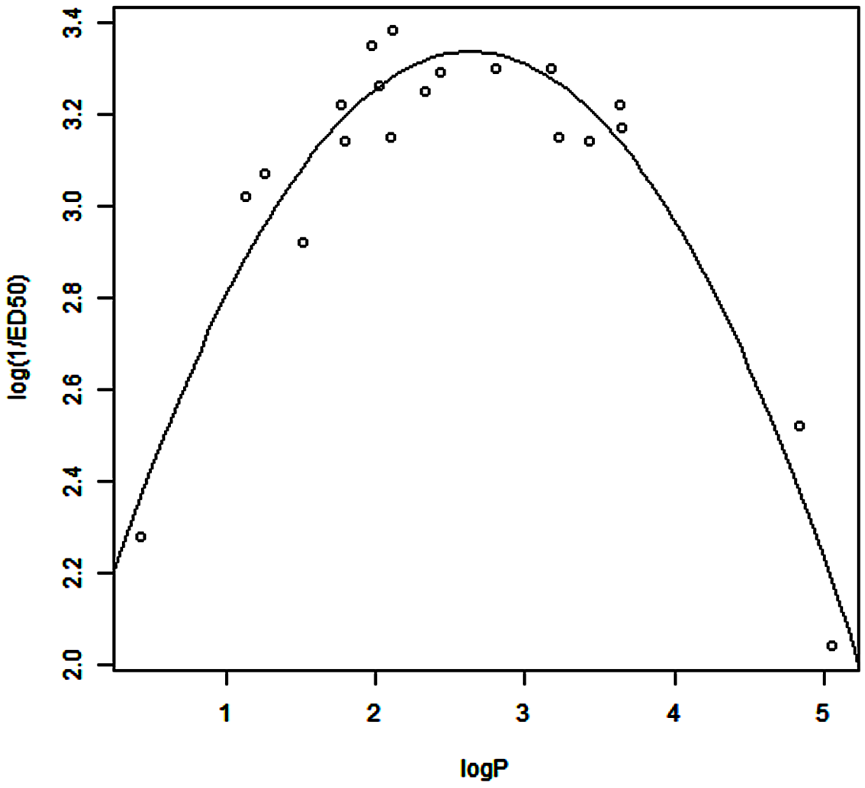
$$\log (1/ED_{50}) = 1.958 + 1.029 \log P - 0.195 (\log P)^2 \quad (15)$$

$n = 20\ r^2 = 0.904\ s = 0.118$

In the 1970s, the pharmaceutical industry was starting to use QSAR in drug design to a significant extent, and that has carried on (Perkins, Fang, Tong, & Welsh, 2003). There is, however, a story (possibly apocryphal) that medicinal chemists, on finding that increasing the hydrophobicity of drug candidates did not always increase potency, would cry despairingly: "Methyl, ethyl, butyl, futile".

Others were sceptical of the fact that some QSAR models incorporated as many as five or six descriptors, and would claim: "Give me five descriptors and I'll draw you an elephant!" This stimulated Wei (1975) to see how many descriptors actually were needed to fit an elephant. He found that 30 descriptors were required before a shape that could be described as elephant-like was obtained; five descriptors gave merely an egg shape.

Figure 4. Anti-inflammatory activity of aspirin derivatives vs. log P



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 1. European QSAR Symposia

Year	Location	Organiser(s)	Symposium proceedings
1973	Prague, Czechoslovakia	Miloň Tichý	Tichý, 1976
1976	Suhl, East Germany	Rainer Franke, Peter Oehme	Franke, & Oehme, 1978
1979	Budapest, Hungary	Ferenc Darvas	Knoll, & Darvas, 1980
1982	Bath, UK	John Dearden	Dearden, 1983
1984	Bad Segeberg, West Germany	Joachim Seydel	Seydel, 1985
1984	Prague, Czechoslovakia[a]	Miloň Tichý	Tichý, 1985
1986	Portorož, Yugoslavia	Dušan Hadži, Borka Jerman-Blažič	Hadži, & Jerman-Blažič, 1987
1988	Interlaken, Switzerland	Jean-Luc Fauchère	Fauchère, 1989
1990	Sorrento, Italy	Carlo Silipo, Antonio Vittoria	Silipo, & Vittoria, 1991
1992	Strasbourg, France	Camille Wermuth	Wermuth, 1993
1994	Barcelona, Spain	Ferran Sanz	Sanz, Giraldo, & Manaut, 1985
1996	Lausanne, Switzerland	Han van de Waterbeemd Bernard Testa Gerd Folkers	van de Waterbeemd, Testa & Folkers, 1997
1998	Copenhagen, Denmark	Klaus Gundertofte	Gundertofte, & Jørgensen, 1999
2000	Düsseldorf, Germany	Hans-Dieter H? ltje	H? ltje, & Sippl, 2001
2002	Bournemouth, UK	Martin Ford, David Livingstone John Dearden Han van de Waterbeemd	Ford, Livingstone, Dearden, & van de Waterbeemd, 2003
2004	Istanbul, Turkey	Esin Aki-Sener	Aki-Sener, & Yalcin, 2006
2006	Mediterranean cruise	Gabriele Cruciani	None
2008	Uppsala, Sweden	Torbjörn Lundstedt	In *QSAR & Combinatorial Science* 28 (2009), no. 8
2010	Rhodes, Greece	Anna Tsantili-Kakoulidou, Dimitris Agrafiotis	In *Molecular Informatics 30* (2011), nos. 2 & 3
2012	Vienna, Austria	Gerhard Ecker	In *Molecular Informatics 32* (2013), nos. 5-6
2014	St. Petersburg, Russia	Vladimir Poroikov	In *Molecular Informatics 33* (2015), nos. 6-7

[a]The Prague meeting was run immediately following the Bad Segeberg meeting, so as to allow delegates to attend both meetings.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

With the increasing interest in QSAR in the 1970s, two series of QSAR conferences were initiated. The first was the European Symposia on QSAR, started by Miloň Tichý in Prague in 1973; they were initially held triennially, but from 1982 have been held biennially. Table 1 gives details of those held to date, and the proceedings arising from them.

The second series was the Gordon Conferences on QSAR, held biennially under the auspices of the Gordon Research Conference organisation (Gordon Research Conferences website). Although Gordon Research Conferences started in 1931, the first QSAR conference was held in 1975, chaired by Paul Craig. Table 2 gives details of those held to date. It should be noted that no publications arise from Gordon Conferences, and no-one is permitted to quote anything presented there. From 2002 The Gordon Conference on QSAR changed its name to the Gordon Conference on Computer-Aided Molecular Design to reflect the broadening scope of *in silico* methods in the design of bioactive compounds.

Increasing interest in QSAR led Joachim (Jo) Seydel to start the first journal dedicated to QSAR in 1982; it was called simply *Quantitative Structure-Activity Relationships*, and Seydel was its editor until 1998. For many years it incorporated abstracts (collected and edited by Ferenc Darvas) of recently published QSAR papers from many different journals, which increased its value to the QSAR community.

Table 2. Gordon Research Conferences on QSAR

Year	Location	Organiser
1975	Franklin Pierce University, NH, USA	Paul Craig
1977	Plymouth State College, NH, USA	Yvonne Martin
1979	Plymouth State College, NH, USA	John Topliss
1981	Plymouth State College, NH, USA	Albert Leo
1983	Plymouth State College, NH, USA	Stefan Unger
1985	Plymouth State College, NH, USA	Richard Cramer
1987	Plymouth State College, NH, USA	Anton Hopfinger
1989	Tilton School, NH, USA	William Dunn
1991	Tilton School, NH, USA	Jeffrey Blaney
1993	Tilton School, NH, USA	Hugo Kubinyi
1995	Tilton School, NH, USA	Robert Pearlman
1997	Tilton School, NH, USA	Herschel Weintraub
1999	Tilton School, NH, USA	Gerald Maggiora
2001	Tilton School, NH, USA	Kate Holloway
2003*	Tilton School, NH, USA	John Van Drie
2005	Tilton School, NH, USA	Peter Jurs
2007	Tilton School, NH, USA	Richard Lewis
2009	Tilton School, NH, USA	Brian Shoichet
2011	Mount Snow, NH, USA	Gerhard Klebe
2013	Mount Snow, NH, USA	Martin Stahl
2015	Mount Snow, NH, USA	Anthony Nicholls

*From 2003, the conference title was changed to Computer Aided Drug Design

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

TRANSPORT THROUGH AN ORGANISM

The question arises as to why activity first increases and then decreases as hydrophobicity increases. Put simply, a chemical with a low partition coefficient will move only slowly from water to lipid, and therefore will arrive slowly at a receptor site after its random walk through a number of lipid membranes. Conversely, a chemical with a high partition coefficient will move quickly from water to lipid, but will leave lipid for the next aqueous phase slowly, therefore again arriving slowly at the receptor site. So the chemical with the most rapid movement will be one with intermediate values of both water→lipid and lipid→water partition coefficient. There are a number of other possible reasons for non-rectilinearity, such as different distribution in different compartments of an organism, steric restrictions at a receptor site, and increased metabolism of more hydrophobic xenobiotics as the organism seeks to excrete them (Kubinyi, 1993).

Penniston et al. (1969) devised a computer-based model of partition-based transport through 20 alternating aqueous and lipid compartments to demonstrate this. Partition coefficient is an equilibrium constant, and can be defined as the ratio of the rate constant of transfer from water to lipid and that of transfer from lipid to water:

$$P = k_{wo}/k_{ow} \tag{16}$$

By setting up differential equations for each compartment, and following the "administration" of a single dose into the first compartment at time = 0, they were able to show that after a reasonable time, the concentration in the nth compartment rose as P increased, and then fell again, exactly as had been observed for the activity in biological systems.

McFarland (1970) developed a similar model, which was modified by Kubinyi (1977), who developed a bilinear model of activity versus log P rather than the parabolic model proposed by Hansch and Fujita (1964). Kubinyi's bilinear equation is:

$$\log 1/C = a\log P - b(\beta P + 1) + c \tag{17}$$

where a, b, β and c are constants. Kubinyi reported numerous examples where his bilinear model gave better fits than did the parabolic model; for example, for the narcotic action of 10 alcohols on the frog ventricle, the bilinear model yielded $R^2 = 0.996$, $s = 0.133$, whilst the parabolic model gave $R^2 = 0.956$, $s = 0.458$. The value of the Kubinyi equation is that it can accommodate different slopes of the upward and downward legs of a biphasic curve; a disadvantage is that it is not simple to use, as the β value has to be obtained by iteration.

Dearden and Townend (1978) also extended the Penniston model, and showed that both duration of action and time to maximal response varied parabolically with hydrophobicity, in each case going through a minimum at some intermediate value of log P. Also, although the Penniston model produced a parabolic type curve of activity against hydrophobicity at a given time after dosage, as had been found to exist in practice (Hansch & Fujita, 1964), Dearden and Townend (1983) showed that, at time to maximal response, the plot of activity against hydrophobicity initially rose, then levelled out instead of falling. This suggests that hydrophobic drugs could be significantly more potent than is indicated by measurement of potency at a fixed time after dosage.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

QSAR IN ENVIRONMENTAL SCIENCES

As shown above, much of the very early work on the relationship between chemical structure and biological activity involved toxicity rather than pharmacological activity. It is therefore somewhat surprising that, following the key paper by Hansch et al. (1962), it was in pharmacology that QSAR was most widely used. In fact, it was not until the early 1980s that QSAR began to be used to any degree in environmental sciences.

One of the earliest environmental QSAR studies (Neely, Branson, & Blau, 1974) found a good rectilinear correlation between bioconcentration factor (BCF) and partition coefficient for some organic environmental pollutants:

$$\log BCF = 0.124 + 0.542 \log P \tag{18}$$

$$n = 8\ r^2 = 0.899\ s = 0.342$$

where BCF = ratio of steady-state concentration of chemical in trout muscle to that in water. Hence BCF can be regarded as an organism-water partition coefficient.

An important paper by Könemann (1981) showed that fish toxicity could be correlated well with log P for industrial chemicals acting via non-polar narcosis:

$$\log 1/LC_{50} = -4.87 + 0.871 \log P \tag{19}$$

$$n = 50\ r^2 = 0.976\ s = 0.237$$

where LC_{50} = concentration to kill 50% of guppies in a given time.

The log P values of the 50 chemicals ranged from –1.35 to 5.69, and Könemann commented that he was unable to measure LC_{50} values of chemicals with log P > 6, presumably because of low aqueous solubility.

A significant result of Könemann's work is the excellent straight-line correlation that he obtained, with no sign of the curvature often found with single-dosage tests on animals. The two situations are quite different, with the fish being supplied with a constant "dose" of the test chemical in the surrounding aqueous milieu. Dearden and Townend (unpublished information) found that their computer-based model of distribution of xenobiotics in an organism (Dearden & Townend, 1978) generated the same type of correlation, namely a straight-line correlation with log P when a constant concentration was maintained in compartment 1 of the model.

Further work has shown (Schultz, Lin, & Arnold, 1991; Cronin, 2003) that different mechanisms of action required different QSAR models. Using the toxicities of chemicals to the aquatic ciliate *Tetrahymena pyriformis*, they reported the following QSARs:

$$\textit{Non-polar narcosis}: \log 1/IGC_{50} = 0.74 \log P - 1.86 \tag{20}$$

$$n = 148\ r^2 = 0.96\ s = 0.21\ F = 3341$$

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Polar narcosis: $\log 1/IGC_{50} = 0.59 \log P - 0.94$ (21)

$n = 119\ r^2 = 0.87\ s = 0.26\ F = 806$

Uncouplers: $\log 1/IGC_{50} = 0.40 \log P - 0.19$ (22)

$n = 12\ r^2 = 0.82\ s = 0.25\ F = 52$

Electrophiles: $\log 1/IGC_{50} = 0.60 \log P - 0.33\ E_{LUMO} - 1.00$ (23)

$n = 239\ r^2 = 0.80\ s = 0.34\ F = 476$

where E_{LUMO} = energy of the lowest unoccupied molecular orbital.

It should be noted that the rectilinear correlations of aquatic toxicity with log P are obtained when the system has reached steady state conditions; typically an aquatic toxicity test is conducted over several days – a 96-hour test period is often used.

The correlations of equations similar to those of equations 20 – 22 converge at log P values of 5-6 (Dearden, 2002).

Following increasing interest in environmental QSAR in the 1980s, a series of international workshops on QSAR in the environmental sciences was started in 1983 by Klaus Kaiser. As with the European QSAR symposia, they were initially held triennially, but from 1988 have been held biennially. Table 3 gives details of those held to date, and the publications arising from them. From 2004 their scope was expanded to include health sciences.

The increasing interest in environmental QSAR also encouraged James Devillers to start in 1993 a journal dedicated largely to that field – *SAR and QSAR in Environmental Research*. Devillers has been its editor since its inception. Environmental QSAR studies tended initially to focus largely on aquatic toxicity and hydrophobicity, but gradually extended to other species and other models.

The predictive ability of QSAR models, particularly in environmental fields, opened up the possibility of predicted toxicities and physicochemical properties being used in regulatory matters. In 2002 a meeting of QSAR experts was held in Setúbal, Portugal, to produce a set of QSAR guidelines that would ensure that QSAR predictions used for regulatory purposes were properly validated. Those guidelines were then taken under the aegis of the Organisation for Economic Co-operation and Development (OECD, 2007). They are:

To facilitate the consideration of a (Q)SAR model for regulatory purposes, it should be associated with the following information:

1. *A defined endpoint;*
2. *An unambiguous algorithm;*
3. *A defined domain of applicability;*
4. *Appropriate measures of goodness-of-fit, robustness and predictivity;*
5. *A mechanistic interpretation, if possible.*

These criteria have assumed more importance with the implementation of the REACH (Registration, Evaluation, Authorisation, and Restriction of Chemicals) legislation in 2007 (REACH, 2007).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 3. International workshops on QSAR in environmental sciences

Year	Location	Organiser(s)	Workshop proceedings
1983	Hamilton, Ontario, Canada	Klaus Kaiser	Kaiser, 1984
1986	Hamilton, Ontario, Canada	Klaus Kaiser	Kaiser, 1987
1988	Knoxville, TN, USA	Terry Schultz	Turner et al., 1988
1990	Veldhoven, Netherlands	Joop Hermens	Hermens & Opperhuizen, 1991
1992	Duluth, MN, USA	Gilman Veith	None
1994	Belgirate, Italy	Walter Karcher	In *SAR QSAR Env. Res.* 3 (1995), nos. 3-4 & 4 (1995), nos. 1-4
1996	Elsinore, Denmark	Fei Chen, Gerrit Schüürmann	Chen & Schüürmann, 1997
1998	Baltimore, MD, USA	John Walker	None
2000	Bourgas, Bulgaria	Ovanes Mekenyan	In *SAR QSAR Env. Res.*, 13 (2002), nos. 1-2
2002	Ottawa, Canada	Roger Breton	In QSAR Comb. Sci. 22 (2003), nos. 1-3
2004 (2004)	Liverpool, UK	Mark Cronin, Judith Madden, David Roberts, John Dearden	In *SAR QSAR Env. Res.* 15 nos. 5-6 & 16 (2005), nos. 1-2
2006 (2007)	Lyon, France	James Devillers, Alain Carpy	In *SAR QSAR Env. Res.* 18 nos. 1-4
2008 (2008)	Syracuse, NY, USA	Philip Howard, Jay Tunkel	In *SAR QSAR Env. Res.* 19 nos. 7-8
2010 (2010)	Montreal, Canada	Kannan Krishnan	In *SAR QSAR Env. Res.* 21 nos. 7-8 & 22 (2011), nos. 1-2
2012 (2013)	Tallinn, Estonia	Uko Maran	In *SAR QSAR Env. Res.* 24 nos. 4-6
2014	Milan, Italy	Emilio Benfenati	None

REACH requires information on up to 18 physicochemical properties and up to 19 toxicological properties, depending in annual supply levels, of chemicals manufactured in or imported into the European Union. The relevant physicochemical properties for REACH that are amenable to QSPR prediction are: melting/freezing point, boiling point, relative density, vapour pressure, surface tension, water solubility, 1-octanol–water partition coefficient, flash point, flammability, explosive properties, self-ignition

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

temperature, adsorption/desorption, dissociation constant, viscosity, and air–water partition coefficient (Henry's law constant). Dearden et al. (2013) have discussed in detail the QSPR approaches to each of these. An e-book (Benfenati, 2012) gives guidance on the use of QSAR methods for REACH.

The required toxicological properties for REACH include, *inter alia*, skin and eye irritation, skin sensitisation, aquatic toxicity, acute toxicity, sub-chronic rodent toxicity, mutagenicity, and carcinogenicity. Dearden and Rowe (2015) have discussed the role of artificial neural network (ANN) QSARs in the determination of the required toxicological properties.

A defined applicability domain (AD) is important for predictive purposes. One should not use a QSAR/QSPR model to predict an activity/property of a compound that is not reasonably similar to those used to develop the model, otherwise one would have low confidence in the prediction. Netzeva et al. (2005) have examined a number of methods for the determination of AD, and Roy et al. (2015) have recently proposed a new simple approach for that purpose.

In recent years more QSAR work is focussing on mechanistically based modelling, not necessarily because of the OECD guidelines, but because it is important to know the mechanistic basis of a toxic effect, as an aid to developing an Adverse Outcome Pathway. An AOP is a sequential chain of causally linked events at different levels of biological organisation that lead to an adverse health or ecotoxicological effect. AOPs are the central element of a toxicological knowledge framework being built to support chemical risk assessment based on mechanistic reasoning. For example, skin sensitisation is an adverse outcome that has been studied intensively, and a number of mechanisms have been identified (Enoch, Madden & Cronin, 2008).

Another emerging area is the use of QSAR as part of an integrated or weight-of-evidence *in silico* modelling strategy, involving, say, QSAR, expert systems and read-across (Hewitt, Ellison, Enoch, Madden & Cronin, 2010; Richarz et al., 2013). The use of several independent modelling methods gives increased confidence in the outcome.

Regulatory authorities such as the U.S. Environmental Protection Agency, the U.S. Food and Drug Administration, and the European Commission, all have great involvement in QSAR and related programs. For example, QSAR work at the European Commission's Joint Research Centre at Ispra, Italy, has been reported (Tsakovska, Lessiargska, Netzeva, & Worth, 2008; Mostrag-Szlichtyng, Zaldivar-Comenges, & Worth, 2010).

The OECD has developed a QSAR Toolbox (OECD QSAR Toolbox), which provides software for grouping chemicals into categories and filling gaps in (eco)toxicity data. Despite its name, it does not perform QSAR analyses. It can be downloaded free of charge.

The QSAR approach to prediction can, of course, be used in fields other than pharmacology and tocicology. An interesting example is its use in predicting the quality of Bordeaux wine (Byron & Ashenfelter, 1995). On the reasonable basis that quality is reflected in price, they reported the following QSAR:

$$\log (\text{price}) = 0.0238\ \text{Age} + 0.616\ \text{Temp (GS)} - 0.00386\ \text{RainA-S} + 0.00117\ \text{RainO-M} \quad (24)$$

$$r^2 = 0.83$$

where Age = age of wine (years), Temp(GS) = average temperature during growing season, RainA-S = amount of rain in August and September, and RainO-M = amount of rain during previous October to March. No other statistics were reported, and no intercept was given.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

DESCRIPTORS

As knowledge and expertise in QSAR grew, it became apparent that the simple physicochemical properties used in early post-1962 modelling were not always adequate. In particular, whole-molecule properties became more widely used, allowing modelling of non-congeneric series to be carried out. One of the basic tenets of QSAR modelling is that the compounds used to develop a QSAR should preferably act by the same mechanism of action, otherwise a good model may not be obtained. However, it is not always easy to determine mechanisms, so an assumption was made that compounds within a congeneric series would be more likely to have the same mechanism of action – which is often but not always true. Even within a congeneric series, simple descriptors are not always adequate. For example, the energy of the lowest unoccupied molecular orbital (E_{LUMO}) is frequently used in modelling the aquatic toxicity of electrophiles (see equation 23).

Over the years, many different types of descriptor have become available, and the QSAR practitioner now has many thousands from which to choose. They include simple physicochemical descriptors, those from quantum mechanics (e.g. atomic charge, E_{LUMO}), and those based on molecular topology, such as molecular connectivities (Kier & Hall, 1976, 1986), electrotopological state indices (Hall, Mohney, & Kier, 1991), and graph-theoretic indices (Basak, Niemi & Veith, 1990).

Nowadays descriptors are generally classified as 2D or 3D. 2D descriptors are those that do not depend on molecular conformation; they include most physicochemical descriptors such as log P, molecular weight, and those calculated from graph theory, such as molecular connectivities. 3D descriptors are sensitive to conformation, and include, for example, quantum chemical descriptors, and molecular surface area.

A very interesting set of 5 descriptors are the so-called solvatochromic descriptors, developed by Kamlet et al (1983). They represent polarity, polarizability, hydrogen bond donation, hydrogen bond acceptance, and molecular size, and have been used to model a wide range of both physicochemical and biological properties. It is pertinent to examine here the QSPR for octanol-water partition coefficient (log P) published by Abraham, Chadha, Whiting, and Mitchell (1994):

$$\log P = 0.088 + 0.562\ R - 1.054\ \pi + 0.034\ \alpha - 3.460\ \beta + 3.814\ V_X \qquad (25)$$

$$n = 613\ r^2 = 0.995\ s = 0.116\ F = 23161$$

where R = excess molar refraction (a measure of polarizability), π = polarity/polarizability, α = hydrogen bond donor ability, β = hydrogen bond acceptor ability, and V_X = McGowan characteristic volume. Since the solvatochromic descriptors are approximately auto-scaled, the magnitude or each coefficient indicates the contribution of each descriptor to log P. It can be seen that H-bond acceptor ability and molecular size are the dominant contributors to log P. The former reflects the very strong H-bond donor ability of water, and the latter indicates that more energy is required to create a cavity in water than in octanol, because of strong water-water hydrogen bonding, hence larger molecules favour octanol.

An unusual application of non-molecular descriptors is in ecotoxicology. Although many QSARs that model ecotoxicity use the standard molecular descriptors, Lithner (1989) found that the reproductive toxicity of metal ions to *Daphnia magna* correlated well with the background concentrations of the metal ions in lake water (Lake Superior, USA):

$$\text{Log } (TOX_{rep}) = 0.95 \log (BACKGR) + 1.95 \qquad (26)$$

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

$n = 22\ r^2 = 0.87$

where TOX_{rep} = toxic effect level, and BACKGR = typical background concentration of metal ion in oligotrophic lake water.

Walker et al. (2007) developed QSARs for the growth inhibition by 17 metal ions of *Helianthus annuus* 'Sunspot' (sunflower). Using physico-chemical descriptors, they obtained:

$$\log (1/EC_{50}) = 0.636 + 0.000209\ \rho - 0.00367\ \Delta H_s - 0.242\ (\log K_1) \quad (27)$$

$n = 17\ r^2_{adj} = 0.81\ s = 0.265\ F = 24.1$

where ρ = metal element density, ΔH_s = enthalpy of formation of metal sulfide, and K_1 = stability constant of metal ion with sulfate.

Using background occurrence levels as descriptors, they obtained:

$$\log (1/EC_{50}) = 1.04 - 0.302\ (\log M_{soil}) + 0.000014\ X - 0.000025\ (\text{Land plants}) \quad (28)$$

$n = 17\ r^2 = 0.83\ s = 0.25\ F = 27.6$

where M_{soli} = average metal concentration in soil, X = median elemental composition of soil, and Land plants = mean elemental content in land plants.

Thus the use of background occurrence levels demonstrates their value in ecotoxicity modeling, for they yielded a better correlation than did the use of physico-chemical descriptors. In general, higher background levels mean lower toxicity.

Large compilations of descriptors are now available (e.g. CODESSA, ADAPT, MOE, DRAGON, HYBOT, Molconn-Z).

With so many descriptors now available, how does one choose those that best model the property under investigation? The first step is to eliminate those descriptors with the same value for all compounds, and then to delete one of each pair with high pair-wise collinearity. The rationale for that is that two such descriptors contribute much the same information, and also that collinearity can distort the statistics of a QSAR (Dearden, Cronin, & Kaiser, 2009).

A number of variable selection methods are available (González, Terán, Saíz-Urra, & Teijeira, 2008). The simplest is probably that of step-wise regression, whereby the best single descriptor is selected first, then the next best which, *together with the first*, gives the best two-descriptor model, and so on. The method is fast, but a drawback is that the best n-descriptor model developed in this way is not necessarily the best overall model, since the latter may not incorporate one or more of the descriptors selected by step-wise regression.

Best sub-sets selection eliminates this problem, but is computationally very expensive and time-consuming; it is generally not recommended for pools of more than 20 descriptors.

A very useful variable selection method is the use of genetic algorithms, which imitate genetic selection (Devillers, 1996). It is relatively fast, and has the advantage that it can provide not only the best n-descriptor QSAR, but also the next several best, which allows the selection of a good QSAR model with appropriate (e.g. easily interpretable) descriptors. A similar method is particle swarm optimization, based on the behavior of bird flocking or fish schooling (Hamzeh-Mivehroud, Sokouti, & Dastmalchi, 2015).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

STATISTICS

In order to be of use, a QSAR model must now have appropriate measures of goodness-of-fit, robustness and predictivity (OECD, 2007). Early QSAR work reported few statistics, and occasionally none. Typically QSARs were reported with a correlation coefficient (r) and a standard error of prediction (s) or other measure of error such as root mean square error (RMSE). That gave an indication of goodness of fit, but not of robustness or predictivity. One way to assess robustness is to use an internal validation procedure such as leave-one-out, whereby one compound is removed, and the QSAR is re-developed on the remaining compounds and the correlation coefficient is re-calculated. The removed compound is then returned and a second one removed, and the procedure is repeated, and so on until all compounds have been removed one by one, and an overall correlation coefficient q is calculated. It should be noted that there is still some controversy regarding the value of q, at least so far as its predictive ability is concerned (Golbraikh & Tropsha, 2002; Gramatica, 2007).

Predictivity is vitally important, for one of the key requirements of a QSAR/QSPR model is that it should be able to make reasonably accurate predictions of the requisite activity/property for compounds not used in the development of the model. It is now widely accepted that predictivity of a QSAR should be carried out by means of external validation. That is, the QSAR is used to predict the activities of a test set comprising compounds that are similar to those used in the training set, but were not used in the training set. This is usually performed by splitting a data set into two, a training set (with which the QSAR will be developed), and a test set, typically in an 80%/20% ratio. It is also recommended (Gramatica, 2007) that external validation should not be carried out on QSARs developed using small data sets (< 25 compounds).

Additional useful statistics are: r^2 adjusted for degrees of freedom, which allows comparison of QSARs with different numbers of descriptors; standard errors of the coefficient of each descriptor, which gives an indication of the valid inclusion of a descriptor in a QSAR; and the Fisher statistic (*F*), which gives an indication of the probability that a QSAR is a chance correlation. Dearden et al. (2009) have discussed these statistics in detail.

The value of using standard errors of coefficients is demonstrated by the following QSAR, which models the selectivity of a series of anti-ischaemic tetraalkylbispidines in the rat (Schön, Anton, Brückner, Messinger, Franke, & Gruska, 1998):

$$\log(\text{selectivity}) = 0.37(\pm 0.33)\,MR_1 - 0.010(\pm 0.007)\,(MR_1)^2 + 0.17(\pm 0.10)\,(MR_{3,4})$$

$$- 0.0043(\pm 0.002)\,(MR_{3,4})^2 + 0.43(\pm 0.40)\,I2 - 3.03 \qquad (29)$$

$$n = 16\ r^2 = 0.950\ s = 0.194\ F = 38.3$$

where MR_n = molar refractivity of substituent at position n, and I2 = indicator variable for unsaturation in substituent R_1. Molar refractivity is essentially a volume term, but has a polarizability component. It can be seen that the standard errors of the coefficients of several of the descriptors in equation 29 are almost as large as the value of the coefficient itself, which casts doubt on the validity of those descriptors. In addition, the QSAR does not comply with the Topliss and Costello rule (1972) (*vide infra*).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

APPROACHES TO QSAR/QSPR MODELLING

With the availability of so many descriptors, and because *a priori* one often does not know which descriptor(s) will give a good model, many QSAR practitioners start with a large pool of descriptors, and use various techniques, such as eliminating those with high pair-wise collinearities, to reduce the number. Nevertheless it is still necessary to select those few descriptors that will give the best model, and to employ techniques such as step-wise regression, best sub-sets selection or genetic algorithms to do so.

Topliss and Costello (1972) showed that if the ratio of the number of compounds in the training set used to develop a QSAR to the number of descriptors in the QSAR was less than five, there was a significant risk that the correlation was by chance and was not valid. Topliss and Edwards (1979) later showed that, if descriptors were selected from a very large pool, that also increased the risk that chance correlations could arise. It is a matter of concern that both guidelines are still often ignored.

Guidelines for the development of QSAR/QSPR models have been published (Walker, Jaworska, Comber, Schultz, & Dearden, 2003; Walker, Dearden, Schultz, Jaworska, & Comber, 2003; Livingstone, 2004).

In fact, many guidelines for QSAR development and validation are ignored or used incorrectly, which prompted Dearden et al. (2009) to report that they had found 21 different types of error in published QSAR papers, including, *inter alia*, incorrect chemical structures, duplication of compounds, errors in descriptor and end-point values, inadequate data, inadequate or incorrect statistics, unacknowledged omission of data points, over-fitting of data, and incorrect validation.

Other QSAR Approaches

The relatively simple approach in which a set of biological activities is correlated with one or more descriptors is termed multiple linear regression (MLR). It is still widely used, but has a number of drawbacks:

1. It cannot handle non-rectilinear relationships between biological activity;
2. It is very sensitive to outliers;
3. The descriptors should be independent of each other.

Hence, over the years numerous other modelling techniques have been employed in QSAR/QSPR analysis. Many of these were developed originally for purposes other than QSAR analysis.

Artificial neural networks (ANNs), so-called because they are conceptually similar to neural networks in organisms, are fairly extensively used in QSAR/QSPR modelling (Dearden & Rowe, 2015). They have the advantage they can incorporate non-rectilinear functions, such as reciprocal and squared terms. Their disadvantages are that they can easily be over-trained, which would reduce their predictivity, and that they do not yield a QSAR/QSPR equation.

Radial basis function networks are a variant of ANNs. Modarresi, Modarress, and Dearden (2007) used them to develop a QSPR model for Henry's law constant (air-water partition coefficient).

Fuzzy ARTMAP is a synthesis of adaptive resonance theory (ART) neural networks and fuzzy logic. An example of its use in QSAR was for the prediction of aqueous solubility of organic compounds (Yaffe, Cohen, Espinosa, Arenas, & Giralt, 2001). It should be noted that in that work the standard error of prediction was given as 0.08 log unit, whereas the standard error of measurement of aqueous solubility

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

is about 0.6 log unit. Livingstone (1995) has pointed out that in QSAR modeling, if the standard error of prediction is less than that of measurement, over-fitting of the model has taken place.

Principal components analysis (PCA) is a technique that uses an orthogonal transformation to convert a set of values of possibly correlated variables (descriptors) into a set of values of uncorrelated variables called principal components. The transformation is defined so that the first PC has the largest possible variance (i.e. accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to the preceding components. It is thus useful in effectively reducing a large descriptor pool to manageable proportions.

Discriminant analysis is one technique that can be used for categorical or classification data, for example active/inactive (1/0), or strong/moderate/weak/inactive. In effect it places a multi-dimensional hyperplane between classes, although it is often found that class discrimination is less than 100%. It may be thought of as semi-quantitative, but since the descriptor values used are quantitative, it is classed as a QSAR/QSPR technique.

Logistic regression is similar to discriminant analysis; it estimates probabilities using a logistic function. Worth and Cronin (2003) have discussed the use of both techniques in the development of classification models for human health effects.

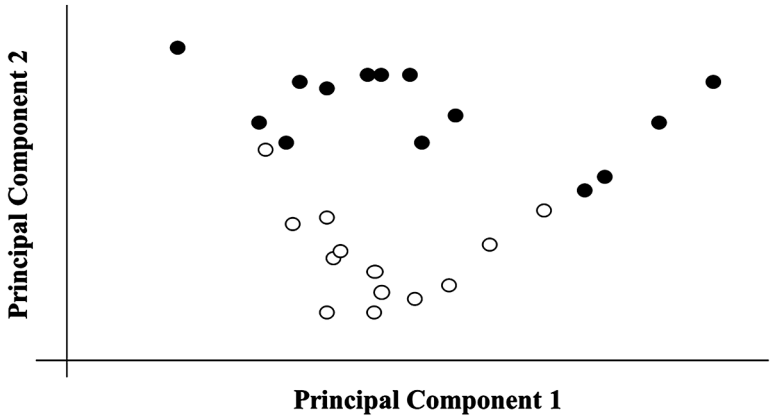
Barratt (1995) utilised both PCA and discriminant analysis in a study of the skin corrosivity of organic acids. From the descriptors log P, molecular volume, melting point and pKa, he obtained two PCs that discriminated between corrosive and non-corrosive organic acids (Figure 5).

Correspondence analysis is conceptually similar to principal components analysis, but applies to categorical rather than continuous data. Cronin and Dearden (1997) used it to determine the skin sensitisation potential of organic compounds.

Support vector machines (SVMs) are supervised learning models with associated learning algorithms that recognise patterns for classification and regression analysis. They were utilised in the development of QSAR models for predicting anti-HIV-1 activity (Darnag, Mazouz, Schmitzer, Villemin, Jarid, & Cherqaoui, 2010). They allow for more convoluted hyperplane boundaries than does simple discriminant analysis.

Adaptive least squares uses a single discriminant function to make decisions for ordered group discrimination. It has the advantage that it can simultaneously consider any number of classes, and thus can

Figure 5. Discrimination of skin corrosive (○) and non-corrosive (●) organic acids (after Barratt, 1995)



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

be used for both classification and regression; it is also useful in dealing with data involving activities at only one or two concentrations, when an ED_{50} or LD_{50} cannot be calculated. Schaper and Saxena (1991) have given several examples of its promising performance in regression analysis.

Cluster analysis is used to classify data into groups or categories, based on chemical and/or biological properties, prior to QSAR analysis. It can therefore be used as a data-reduction tool. It is useful when a data set includes chemicals with different mechanisms of action. Yuan and Parrill (2005) used it in the 3D-QSAR analysis of HIV-1 integrase inhibitors.

Factor analysis is another data-reduction method, which can also be used for variable selection. Roy and Ghosh (2004) used it for the latter purpose in the modeling of the toxicity of phenylsulfonyl carboxylates to *Vibrio fischeri*.

Partial least squares analysis (PLS) extracts orthogonal (non-correlated) linear combinations of descriptors that explain variances in both descriptors and responses. It is useful when x-variables (descriptors) are correlated, when y-variables (responses) are correlated, and when the number of x-variables greatly exceeds the number of y-variables; this last is frequently the case in modern QSAR analysis, because of the large number of molecular descriptors now available to QSAR practitioners. The method has been well described by, *inter alia*, its developer, H. Wold (Wold, Ruhe, Wold, & Dunn, 1984).

Canonical correlation analysis (CCA) can be used if there are correlations among variables, both dependent (responses) and independent (descriptors). It finds linear combinations of variables with maximum correlation, and can be considered the equivalent of PCA and PLS analysis. Livingstone et al. (2001) used it to predict simultaneously both aqueous solubilities and octanol-water partition coefficients of organic chemicals.

The k-nearest neighbours (k-NN) algorithm is a non-parametric method (i.e. descriptors are not required) used for both classification and regression. The input consists of the (k) closest training set examples. In regression, the output is the property value of the query compound as an average of the values of its k nearest neighbours. Raevsky et al. (2014) used a k-NN approach to model aqueous solubilities of a large number of unionised organic compounds.

Classification and regression trees (CART) is a hierarchical method that determines a consecutive set of *if-then* conditions that allow accurate prediction or classification. Graphically the method resembles a tree, with branching at the *if-then* nodes. It has a number of attractive features, including simplicity, interpretability, capacity to handle large data sets and model non-linearities, no assumptions regarding data distribution, immunity to outliers and collinearities. Tan et al. (2010) have used the method to model the antimicrobial potencies of diverse agents against *Candida albicans*.

Random forests are an extension of CART; they are ensemble learning methods that use a number of decision trees, the output being the mode of the classes in classification modeling, and the mean prediction of the individual trees in regression modeling. Random forests correct for the overfitting that can occur with decision trees. An example of their use in QSAR is the prediction of toxicity of diverse organic chemicals to the aquatic ciliate *Tetrahymena pyriformis* (Polishchuk, Muratov, Artemenko, Kolumbin, Muratove, & Kuz'min, 2009).

SIMCA (Soft Independent Modeling of Class Analogy) was developed to enhance the classification potential of principal component models. It constructs separate models for different classes, and new objects (chemicals) of unknown class can be projected into the same multivariate space as the training set. It has the advantage of being able to handle large numbers of collinear variables. Giaginis et al. (2014) used it for multivariate data analysis in the QSAR modeling of redistribution of antidepressants.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Mapping techniques such as Kohonen mapping and non-linear mapping are visualization methods; they map a set of vectorial samples onto a two-dimensional lattice whilst preserving the topology of the original space. The resulting patterns can serve as indicators of structure-activity relationships, and can also aid in classification. Polanski et al. (2000) used Kohonen mapping to aid in the selection of new artificial sweetener candidates. Non-linear mapping in QSAR and QSPR has been reviewed by Domine et al. (1993).

Consensus modelling is an approach used in both drug research and toxicology; it involves taking an account, for example by averaging, of predictions from a number of QSAR and/or other models. Matthews et al. (2008) combined predictions from five software programs to improve the prediction of chemical carcinogenesis in rodents. Mitra et al. (2012) used consensus modelling to predict drug-induced adverse reactions.

An important 3D-QSAR approach is comparative molecular field analysis (CoMFA), developed by Cramer et al. (1988). It allows comparison of molecules by aligning them in space and mapping their steric and electrostatic fields on a 3-D grid. It overcomes the usual QSAR problem of an excessive number of descriptors (molecular field values) by the use of partial least squares statistics. CoMFA requires molecular alignment, which can be difficult, and the electrostatic potentials used are very steep at the van der Waals surface, meaning that the potential energy changes rapidly. A variation of CoMFA is CoMSIA (comparative molecular similarity index analysis), developed by Klebe and Abraham (1999), which includes hydrogen bonding as well as steric and electrostatic fields. The contour maps obtained are better than those from CoMFA, in that they are easy to interpret and very intuitive as a visualization tool. da Cunha et al. (2009) used both CoMFA and CoMSIA to generate 3D-QSAR predictions of the activity of some HIV-1 protease inhibitors. Yuan and Parrill (2005) used the related technique of molecular field analysis (MFA) to model the activity of a series of HIV-1 integrase inhibitors.

A similar approach is GRID, which uses a smoother potential function than the Lennard-Jones type used by CoMFA. It explores not only steric and electrostatic potentials, but also hydrogen bonding potential, using a variety of probes. Kim et al. (1993) found it to be superior to CoMFA in a study of benzodiazepines.

Molecular shape analysis involves the characterization and use of molecular shape in the development of a QSAR model (Roy, Kar, & Das, 2015b). It has the advantage that knowledge of receptor geometry is not required. Rowberg et al. (1994) used it in the QSAR modelling of 1-(phenylcarbamoyl)-2-pyrazolin insecticides. Comparative receptor surface analysis (CoRSA) is another technique that does not require knowledge of receptor geometry. It generates a virtual receptor represented as points on a surface. The CoRSA descriptors are then used in, for example, a PLS analysis to construct a QSAR model. It was used by Ivanciuc et al. (2001) to model the calcium channel antagonist activity of some dihydropyridine derivatives.

Common reactivity pattern analysis (COREPA) uses a Bayesian probabilistic method to identify common structural characteristics among chemicals that elicit similar biological activity or class in a context that allows many possible conformations of individual chemicals and the probability distribution of molecular descriptor values instead of single parameter values for each chemical. The common structural characteristics can then be encoded into a decision tree to screen large and structurally heterogeneous chemical libraries for the required biological activity (Mekenyan et al., 1999).

In the fragment-based QSAR (FB-QSAR) approach the bioactivities of molecules are correlated with the physicochemical properties of molecular fragments through two sets of coefficients in linear free energy equations. One coefficient set is for the physico-chemical properties and the other for the weight

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

factors of the molecular fragments. Then an iterative double least square technique is developed to solve the two sets of coefficients in a training data set alternately and iteratively. The FB-QSAR approach can remarkably enhance the predictive power and provide more structural insights into rational drug design. Du et al. (2009) used it to build a predictive model of neuraminidase inhibitors for drug development against H5N1 influenza virus. Group-based QSAR (G-QSAR) is an extrapolation of FB-QSAR that allows substituent interactions as fragment-specific descriptors to account for fragment interactions (Ajmani, Jadhav, & Kulkarni, 2009). It has the advantage that it highlights the molecular site(s) where optimization is required in drug design. This means that it can deal with the inverse QSAR problem; that is, it offers a method for the design of new molecules from the QSAR model. There have, of course, been a number of other studies of inverse QSAR (e.g. Wong & Burkowski, 2009; Hasegawa, Kimura & Funatsu, 2009).

Another important fragment-based QSAR technique that does not require molecular alignment is hologram-based QSAR (HQSAR). In this approach fragments are assigned integer values, and these are then used to make a fixed-length integer array, termed a molecular hologram; the bin occupancies of the hologram are used as descriptors. The approach encodes all possible fragments. PLS is then used to build a QSAR model. In a comparative study of CoMFA, CoMSIA and HQSAR in modeling acetylcholinesterase inhibitors, Jiang et al. (2013) found HQSAR superior to the other techniques based on internal (q^2) validation, but intermediate based on external validation.

Despite the vast number of published papers, edited books and book chapters on QSAR, there are but a few books that can be called QSAR text-books, with basic information and practical advice on how to develop, validate and use QSARs. The first such book was by Yvonne Martin (1978), followed over 30 years later by a second edition (Martin, 2010). Others were by Raevsky (1984) (published in Russian), and Kubinyi (1993); the latest are those of Roy et al. (2015a, 2015b).

SOME NEWER APPLICATIONS OF QSAR

Recent QSAR developments include the use of higher dimensional QSAR modeling, so it is appropriate to consider what dimensionality means in QSAR terms. Greener (2007) gave the following definitions:

1D-QSAR: Activity modeled by simple physico-chemical properties such as log P;
2D-QSAR: Activity modeled by structural patterns such as molecular connectivities;
3D-QSAR: Activity modeled by 3D structure. In receptor modeling, the QSAR is based on the interactions of ligands with a 3D receptor surrogate;
4D-QSAR: As with 3D, but with multiple representations of ligand conformation/orientation;
5D-QSAR: As with 4D, but with multiple representations of induced fit scenarios;
6D-QSAR: As with 5D, but with multiple representations of solvation models.

Several 4D-QSAR approaches have been developed. LQTA-QSAR (Martins et al., 2009) calculates intermolecular interaction energies at each grid point, considering all aligned conformations from molecular dynamics calculations. SOM-4D-QSAR uses a self-organizing map approach to classify data. Multiple conformations of ligands are obtained, and typically an active bound conformation is searched, whilst taking conformational flexibility into account. Bak et al. (2014) used the method successfully to study the structure-affinity relationship of a set of anthraquinone dyes adsorbed on cellulose fiber.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Receptor-independent 4D-QSAR (RI 4D-QSAR) can be used when the pharmacophoric features of a ligand are needed. Senese and Hopfinger (2003) obtained good predictions of the inhibitory activity of norstatine derivatives against HIV-1protease with RI 4D-QSAR; their models also correctly identified the hydrophobic nature of the HIV-1 protease receptor site. Receptor-dependent (RD) 4D-QSAR can be used when 3D structures of relevant macromolecules are available, which can prediction of the binding site. The main aim is to map the ligand-receptor interaction. da Rocha Pita et al. (2012), studying peptidemimetic inhibitors of *Trypanosoma cruzi* tryptanothione reductase, evaluated multiple ligand conformations from molecular dynamics simulations. They were also able to identify some tryptanothione reductase subsites for further development of the drug discovery process.

In 5D-QSAR, in addition to allowing for multiple conformations, orientations and protonation states of ligand molecules, an extra degree of freedom – the fifth dimension – allows for a multiple representation of the topology of the surrogate receptor. Vedani and Dobler (2002) used the technique

In 6D-QSAR a further dimension – the solvation function – is added to improve the QSAR analysis. The solvation expression is either ligand desolvation or solvent stripping when ligand and receptor bind, and includes a range of binding forces, including hydrogen bonding. Peristera et al. (2009) used the technique to determine the binding mode of a series of steroid and non-steroid compounds to the human mineralocorticoid receptor, then used the data to predict the binding affinities of a series of anabolic steroids. Damale et al. (2014) have recently reviewed recent advances in 4D-6D QSAR.The latest extension to the nD-QSAR progression is 7D-QSAR, developed by Polanski (2009). It incorporates real or target-based receptor model data. It may be noted that Wade is quoted (Greener, 2007) as saying that the "dimensional" terminology is misleading, because additional Ds above 3D are not dimensions in the accepted meaning of the word.

Pharmacophore mapping is a technique used when active molecules are known but the target receptor structure is not available (Leach & Gillet, 2007). John et al. (2011) used it in the discovery of new human renin inhibitors; with a training set of 18 active and inactive compounds, and a test set of 93 compounds, they were able to define the main receptor characteristics of the renin receptor and design several novel inhibitors.

Homology mapping uses known amino acid sequences to build a model receptor that will accommodate ligands known to be active against a given unknown receptor (Vyas, Ukawala, Ghate & Chintha, 2012). Tuccinardi et al. (2007) used homology mapping to construct a model of the carbonic anhydrase IX receptor. They then used the best poses of a large number of known inhibitors as an alignment tool to develop the 3D-QSAR model. A similar study on the CYP11B1 receptor was made by Yu et al. (2015).

When no ligands for a given receptor are known, the technique of de novo ligand design may be used. A known target structure is required, and software is used to generate possible ligands by docking fragments into potential target binding sites. These fragments are then joined to an anchor point of a partial ligand in the binding cavity. The fit with the receptor is then determined and the docking scored. Schneider and Fechner (2005) have reviewed the approach. Its use is exemplified by Rogers-Evans et al. (2004), who employed it to identify novel cannabinoid receptor ligands.

A related approach is structure-based ligand design (Gubernator & Bohm, 1998). It requires a crystal structure, or a model derived from the crystal structure of a closely related homolog, of a target protein, preferably complexed with a ligand. This complex can provide the binding mode and conformation of the ligand, and indicates essential aspects of its binding affinity. This information can then be used to develop new ligands, or for queries for virtual screening of large compound libraries. Kroemer (2007)

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

has discussed the methods and software available for structure-based ligand design, and the challenges remaining.

Another related technique is molecular docking, which preferably requires the protein binding site to be known, in order to increase docking efficiency (Meng, Zhang, Mezei, & Cui, 2011). It is often used with a flexible ligand and rigid receptor, but more recently efforts have been made to deal with receptor flexibility, although that poses major challenges. B-Rao et al. (2009) have discussed the problems of receptor flexibility in docking. Like de novo and structure-based molecular design, the method is valuable for the design of potential new drug candidates.

A very important emerging area is that of nanomaterials, which are being increasingly used in a wide range of applications. A main concern is potential toxicity; the small size of nanoparticles (typically 1 – 100 nm) means that they can readily be absorbed by organisms. Currently there is a great paucity of toxicity data in this area (Richarz et al., 2015), and a global effort is needed to provide and evaluate such data. The specific characteristics of nanoparticles, such as shape, surface charge, sate of dispersion/aggregation, particle size distribution, solubility and porosity can all contribute to their properties, which can have both positive and negative effects.

Many nanomaterials are metal-containing chemicals, e.g. metal oxides. Puzyn et al. (2011) developed a QSAR model to predict accurately the toxicity of 17 metal oxide nanoparticles to *Escherichia coli*. Burello and Worth (2011) developed a theoretical framework for predicting the oxidative stress potential of oxide nanoparticles, and have developed a structure-reactivity model for that purpose (Burello & Worth, 2012).

Singh and Gupta (2014) have developed good QSAR models using decision tree approaches for predicting the activity of a range of nanomaterials, such as surface-modified carbon nanotubes and fullerene derivatives, as well as for iron oxide nanoparticles cross-linked with amine groups and with different small molecules attached. Other QSAR modeling techniques have also been used by different workers (Singh & Gupta, 2014), and Puzyn et al. (2010) have given a good survey of the field Oksel et al. (2015) have critically reviewed QSAR modeling of nanomaterial toxicity.

A relatively new addition to QSAR modeling is matched molecular pair analysis (MMPA). A MMP is a pair of molecules that differ only in a single minor point change, and noting the effect of that change can help in understanding mechanisms of action. Leach et al. (2006) used MMPs as a guide in the optimization of pharmaceutical properties, both physico-chemical and biological. Sushko et al. (2014) found that MMPA allowed the identification of large numbers of interesting transformations that could be used to drive the hit-to-lead optimization process.

Most QSAR analyses are carried out using tests carried out on single chemicals. However, most medicines and most toxicants are mixtures, which may or may not behave additively.

QSAR studies on mixture toxicity are not new, although progress has been slow. In a review article, Muratov et al. (2012) commented that existing efforts could be considered as a foundation for future work in the field. They suggest that the use of non-additive mixture descriptors, in combination with best modern practices, will significantly improve the quality of QSAR studies of mixtures. Muratov et al. (2014) used an "everything out" validation approach in QSAR analysis of binary mixtures, whereby external validation sets are formed of all binary combinations of two compounds excluded from the training set, and found it to give lower prediction errors than "compound out" validation.

A potentially valuable approach to QSAR analysis of drug-drug interactions (DDIs) was proposed by Bulusu et al. (2015). Predictive models based on about 25,000 DDIs caused by more than 200 drugs

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

were developed for inhibition of four major cytochrome P450 isoforms, and were validated using a "compounds out" strategy specially developed for validation of QSAR models of mixtures.

With increasing land pollution, it is anticipated that more QSAR modeling of ecotoxicity will be required. The Japanese National Institute for Environmental Studies used QSAR in the development of the Kashinhou tool for ecotoxicity (Furuhama et al., 2010). The ERICA model (Boriani, Benfenati, Baderna, & Thomsen, 2013), developed at the Mario Negri Institute, Milan, Italy, assesses the effect of xenobiotics introduced into the environment. It is claimed that it will be useful for stakeholders such as chemical companies to predict the effects that a new chemical could have on the environment. It is also anticipated, now that the European Union REACH legislation is fully in place, that there will be even more QSAR modeling of toxicity, and especially of chronic toxicity, bearing in mind that REACH requires that animal testing be carried out only as a last resort, and is totally banned for cosmetics products and ingredients.

Ionic liquids are becoming increasingly important, for they are powerful solvents, generally miscible with water, and are electrically conducting. A large number are now available, but there is a paucity of toxicity data for them, and QSAR is starting to play an important rôle in their assessment. Lucs et al. (2010) have successfully modeled their toxicity to the marine bacterium *Vibrio fischeri*. Das and Roy (2013) have reviewed recent progress in modelling the design, properties and toxicity of ionic liquids.

Plants have been used as a source of medicines since time immemorial, and even now only a small fraction of the world's plants have been tested for the medicinal properties of their constituent chemicals. Their medicinal potential is therefore vast, and modern QSAR methods are now being applied to search for new phytochemicals as drugs and other chemical uses. For example, the US Food and Drug Administration uses QSAR expert systems to predict the carcinogenic potential of phytochemicals (Valerio, Kruhlak, & Benz, 2010). Kar and Roy (2012) have reviewed the field.

For further details of the above and other newer QSAR approaches, see publications by Dudek et al. (2006), Verma et al. (2010), and Roy et al. (2015b).

QSAR SOFTWARE

Probably the first QSAR software was the ClogP program for the calculation of log P values, developed at Pomona College (Hansch & Leo, 1979). Since then there have been many log P prediction programs developed, including some pharmaceutical company in-house programs. Mannhold and Dross (1996) tested the predictive capabilities of 14 such programs, using a set of 138 diverse organic compounds. Dearden et al. (2003) updated that study, again testing 14 programs. Dearden has also tested a number of software programs for the prediction of melting point, boiling point and vapour pressure (2003), and of aqueous solubility (2006). Dearden et al. (2007) and Liao and Nicklaus (2009) tested a number of software programs for the prediction of pKa.

In recent years much more QSAR/QSPR software has been developed, for the prediction of a wide range of toxic effects as well as for physicochemical properties. Dearden (2010, 2012) has discussed available software for prediction of toxicity and physicochemical properties. What is still missing, especially for toxicity software, is independent assessment of the comparative predictive performance of the various software packages.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

WHITHER QSAR?

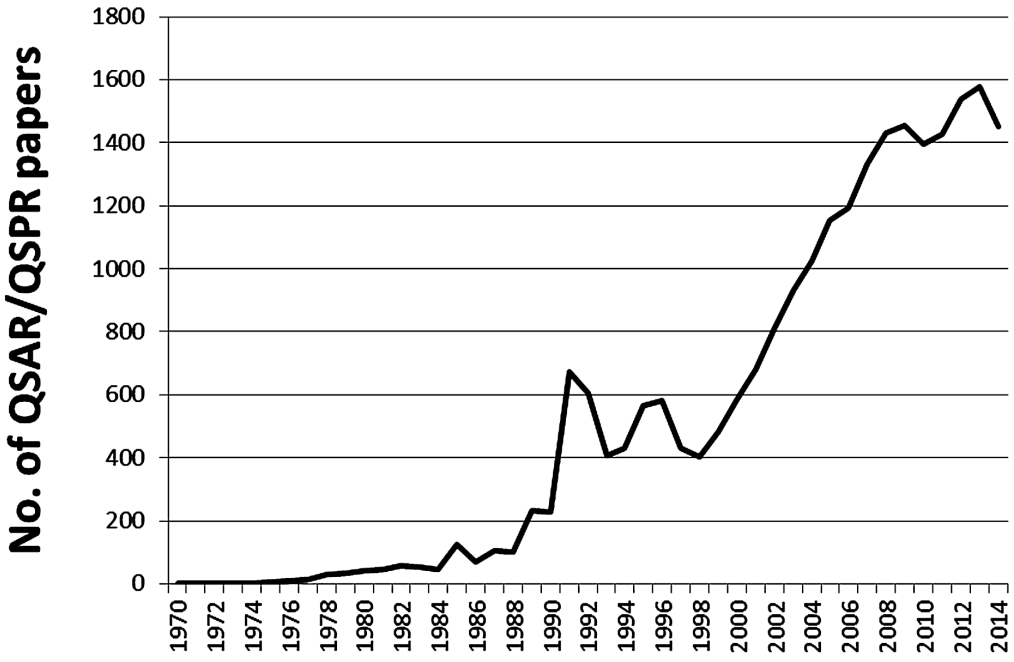
Over the years since 1962, various QSAR practitioners have raised doubts as to the longevity of QSAR. For example, Unger (1984), Dearden (1987), Doweyko (2008), and Combes (2010) have all expressed concern as to whether QSAR would survive for long, but all optimistically concluded that it would do so, albeit perhaps in other guises. Doweyko's (2008) conclusion is: "Thus, QSAR lives on, not only as a stand-alone technique, but even more so in disguised forms within the more popular drug design approaches of the modern era".

One guide as the health of a subject is the number of publications emanating from its practitioners. A search of Web of Science (Web of Science), using the search terms "QSAR", "QSPR", and "quantitative structure-activity relationship" from 1962 onwards yielded the numbers of publications shown in Figure 6. The first paper to use the term "QSAR" appeared in 1973 (Franke, & Oehme, 1973).

Of course, the numbers shown cannot be fully accurate; some QSAR papers do not mention "QSAR", and some non-QSAR papers may do so. However, the trend is clear, if not entirely expected. Why, for example, did publications increase greatly in 1991, only to fall again in 1993, rise again in 1995 and fall again in 1997? Could these variations reflect the increasing use of QSAR in the environmental sciences, and possible variable enthusiasm for QSAR? Is the slight fall in 2014 a sign of things to come, or merely a blip?

An interesting paper by Li et al. (2014) looked at QSAR/QSPR publication rates from 1993 to 2012. Their criteria included the requirement that "QSAR" or "QSPR" had to be mentioned on the first page of a publication, in order to ensure that the focus was definitely on the topic. They also used keywords such as CoMFA. Their numbers were much lower than those found in this work; for example, for 2004

Figure 6. Publication of papers on QSAR and QSPR



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

this work found 1028 publications, whereas Li, Hu, and Ho (2014) found 664. However, over the period 1993-2012, the two data sets correlated very well ($r^2 = 0.969$), meaning that both data sets followed the same trends.

A recent comprehensive overview of QSAR (Cherkasov et al., 2014) gave this optimistic and encouraging affirmation: "QSAR modeling is widely practiced in academy, industry, and government institutions around the world. Recent observations suggest that following years of strong-dominance by the structure-based methods, the value of statistically based QSAR approaches in helping to guide lead optimization is starting to be appreciatively reconsidered by leaders of several larger CADD (computer-aided drug design) groups. QSAR models find broad application for assessing potential impacts of chemicals, materials, and nanomaterials on human health and ecological systems. An area of active QSAR expansion is in the use of predictive models for regulatory purposes by government agencies, where a still growing number of specialized regulatory tools and databases are being developed and validated".

It is apposite to close with a quotation (Millay, 1988) from Edna St. Vincent Millay (1892-1950), an American poet. Al Leo, close associate of Corwin Hansch, used it on his business card:

Upon this gifted age, in its dark hour,
Rains from the sky a meteoric shower
Of facts…they lie unquestioned, uncombined.
Wisdom enough to leech us of our ill
Is daily spun; but there exists no loom
To weave it into fabric.

It is perhaps stretching credibility too far to say that in our gifted age there is no loom to weave myriad facts into patterns. There can, however, be no doubt that QSAR is one of the most powerful of looms, capable of creating wonderful fabrics for our health and well-being.

ACKNOWLEDGMENT

I am grateful to Fabian Steinmetz for help with German texts, and to Claire Mellor and David Ebbrell for help with diagrams and photographs.

REFERENCES

Abraham, M. H., Chadha, H. S., Whiting, G. S., & Mitchell, R. C. (1994). Hydrogen bonding. 32. An analysis of water-octanol and water-alkane partitioning and the Δlog P parameter of Seiler. *Journal of Pharmaceutical Sciences*, *83*(8), 1085–1100. doi:10.1002/jps.2600830806 PMID:7983591

ADAPT. (n. d.). Retrieved from research.chem.psu.edu.pcjgroup/adapt.html

Ajmani, S., Jadhav, K., & Kulkarni, S. A. (2009). Group-based QSAR (G-QSAR): Mitigating interpretation challenges in QSAR. *QSAR & Combinatorial Science*, *28*(1), 36–51. doi:10.1002/qsar.200810063

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Aki-Sener, E., & Yalcin, I. (Eds.). (2006). *QSAR and molecular modelling in rational design of bioactive molecules*. Ankara, Turkey: Başkent Publishers.

B-Rao, C., Subramanian, J., & Sharma, S. D.B-Rao. (2009). Managing protein flexibility in docking and its applications. *Drug Discovery Today*, *14*(7-8), 394–400. doi:10.1016/j.drudis.2009.01.003 PMID:19185058

Babbage, C. (1837). Ninth Bridgewater Treatise. Republished by Cambridge University Press, Cambridge, U.K., 2009.

Bak, A., Wyszomirski, M., Magdziarz, T., Smolinski, A., & Polanski, J. (2014). Structure-based modeling of dye-fiber affinity with SOM-4D-QSAR paradigm: Application to set of anthraquinone derivatives. *Combinatorial Chemistry & High Throughput Screening*, *17*(6), 485–502. doi:10.2174/138620731766 6140205195252 PMID:24499310

Barratt, M. D. (1995). The role of structure-activity relationships and expert systems in alternative strategies for the determination of skin sensitisation, skin corrosivity and eye irritation. *Alternatives to Laboratory Animals*, *23*, 111–122.

Basak, S. C., Niemi, G. J., & Veith, G. D. (1990). A graph-theoretic approach to predicting molecular properties. *Mathematical and Computer Modelling*, *14*, 511–516. doi:10.1016/0895-7177(90)90235-F

Béguyer de Chancourtois, A.-E. (1862). Vis tellurique: classement naturel des corps simples ou radicaux, obtenu au moyen d'un systeme de classification helicoidale et numerique. *Comptes Rendus Hebdomadaire des Séances de l'Academie des Sciences, 54*, 757-761; 840-843; 967-971.

Bell, P. H., & Roblin, R. O. Jr. (1942). Studies in chemotherapy. VII. A theory of the relation of structure to activity of sulfanilamide type compounds. *Journal of the American Chemical Society*, *64*(12), 2905–2917. doi:10.1021/ja01264a055

Benfenati, E. (Ed.). (2012). *Theory, guidance and applications on QSAR and REACH* (e-book). Retrieved from www.orchestra-qsar.eu

Blake, J. (1841). On the action of certain inorganic compounds when introduced directly into the blood. *Proceedings of the Royal Society of London*, *4*(0), 285–286. doi:10.1098/rspl.1837.0144

Blake, J. (1884). On the connection between physiological action and chemical constitution. *The Journal of Physiology*, *5*(1), 36–44. doi:10.1113/jphysiol.1884.sp000148 PMID:16991361

Boriani, E., Benfenati, E., Baderna, D., & Thomsen, M. (2013). Application of ERICA to evaluation of ecosystem health according to sustainability threshold for chemical impact. *The Science of the Total Environment*, *443*, 134–142. doi:10.1016/j.scitotenv.2012.10.025 PMID:23186630

Brown, A., C, , & Fraser, T. R. (1868-69). On the connection between chemical constitution and physiological action; with special reference to the physiological action of the salts of the ammonium bases derived from strychnia, brucia, thebaia, codeia, morphia, and nicotia. *Transactions of the Royal Society of Edinburgh*, *25*(01), 151–203. doi:10.1017/S0080456800028155 PMID:17230757

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Bruice, T. C., Kharasch, N., & Winzler, R. J. (1956). A correlation of thyroxine-like activity and chemical structure. *Archives of Biochemistry and Biophysics*, *62*(2), 305–317. doi:10.1016/0003-9861(56)90129-1 PMID:13328119

Brunton, T. L. (1880). *Pharmacology and therapeutics*. London, UK: Macmillan.

Bulusu, K. C., Guha, R., Mason, D. J., Lewis, R. P. I., Muratov, E., Kalantar Motamedi, Y., & Bender, A. et al. (2015). Modelling of compound combination effects applications to efficacy and toxicity: State-of-the-art, challenges, and perspectives. *Drug Discovery Today*. doi:10.1016/j.drudis.2015.09.003 PMID:26360051

Burello, E., & Worth, A. P. (2011). A theoretical framework for predicting the oxidative stress potential of oxide nanoparticles. *Nanotoxicology*, *5*(2), 228–235. doi:10.3109/17435390.2010.502980 PMID:21609138

Burello, E., & Worth, A. P. (2012). Development and evaluation of structure-reactivity models for predicting the *in vitro* oxidative stress of metal oxide nanoparticles. In J. Leszczynski & T. Puzyn (Eds.), *RSC nanoscience & nanotechnology no. 25. Towards efficient designing of safe nanomaterials: innovative merge of computational approaches and experimental techniques* (pp. 258–284). London, UK: Royal Society of Chemistry. doi:10.1039/9781849735476-00257

Bynum, W. F. (1970). Chemical structure and pharmacological action: A chapter in the history of 19th century molecular pharmacology. *Bulletin of the History of Medicine*, *44*, 518–538. PMID:4925018

Byron, R. P., & Ashenfelter, O. (1995). Predicting the quality of an unborn grange. *The Economic Record*, *71*(1), 40–53. doi:10.1111/j.1475-4932.1995.tb01870.x

Chen, F., & Schüürmann, G. (Eds.). (1997). *Quantitative structure-activity relationships in environmental sciences – VII*. Pensacola, FL: SETAC Press.

Cherkasov, A., Muratov, E. N., Fourches, D., Varnek, A., Baskin, I. I., Cronin, M., & Tropsha, A. et al. (2014). QSAR modeling: Where have you been? Where are you going to? *Journal of Medicinal Chemistry*, *57*(12), 4977–5010. doi:10.1021/jm4004285 PMID:24351051

CODESSA. (n. d.). Retrieved from www.codessa-pro.com

Combes, R. D. (2010). Is computational toxicology withering on the vine? *Archives of Toxicology*, *84*(4), 333–336. doi:10.1007/s00204-010-0528-6 PMID:20204328

Cooke, J.P. (1854). The numerical relation between the atomic weights, with some thoughts on the classification of the chemical elements. *American Journal of Science* (2nd Series), *17*, 387-407.

Cramer, R. D., Patterson, D. E., & Bunce, J. D. (1988). Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *Journal of the American Chemical Society*, *110*(18), 5959–5967. doi:10.1021/ja00226a005 PMID:22148765

Cronin, M. T. D. (2003). Quantitative structure-activity relationships for acute aquatic toxicity: the role of mechanism of toxic action in successful modeling. In R. Benigni (Ed.), *Quantitative structure-activity relationship (QSAR) models of mutagens and carcinogens* (pp. 235–258). Boca Raton, FL: CRC Press. doi:10.1201/9780203010822.ch8

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Cronin, M. T. D., & Dearden, J. C. (1997). Correspondence analysis of the skin sensitization potential of organic chemicals. *Quantitative Structure-Activity Relationships*, *16*(1), 33–37. doi:10.1002/qsar.19970160106

Cros, A. F. A. (1863). Action de l'alcool amylique sur l'organisme. Doctoral thesis, University of Strasbourg, France. In R.L. Lipnick, (1989). Narcosis, electrophile and proelectrophile mechanisms: application of SAR and QSAR. *Environmental Toxicology and Chemistry, 8*, 1-12.

da Cunha, E. F. F., Sippl, W., de Castro Ramalho, T., Ceva Antunes, O. A., de Alencastro, R. B., & Albuquerque, M. G. (2009). 3D-QSAR CoMFA/CoMSIA models based on theoretical active conformers of HOE/BAY-793 analogs derived from HIV-1 protease inhibitor complexes. *European Journal of Medicinal Chemistry*, *44*(11), 4344–4352. doi:10.1016/j.ejmech.2009.05.016 PMID:19616874

da Rocha Pita, S. S., Albuquerque, M. G., Rodrigues, C. R., Castro, H. C., & Hopfinger, A. J. (2012). Receptor-dependent 4D-QSAR analysis of peptidemimetic inhibitors of *Trypanosoma cruzi* tryptanothione reductase with receptor-based alignment. *Chemical Biology & Drug Design*, *79*(5), 740–748. doi:10.1111/j.1747-0285.2012.01338.x PMID:22269140

da Vinci, L. (n. d.). *The Notebooks of Leonardo da Vinci.* Retrieved from http://refspace.com/quotes/Leonardo_da_Vinci/mathematics

Damale, M. G., Harke, S. N., Kalam Khan, F. A., Shinde, D. B., & Sangshetti, J. N. (2014). Recent advances in multidimensional QSAR (4D-6D): A critical review. *Mini Reviews in Medicinal Chemistry*, *14*(1), 35–55. doi:10.2174/1389557513136660104 PMID:24195665

Darnag, R., Mostapha Mazouz, E. L., Schmitzer, A., Villemin, D., Jarid, A., & Cherqaoui, D. (2010). Support vector machines: Development of QSAR models for predicting anti-HIV-1 activities of TIBO derivatives. *European Journal of Medicinal Chemistry*, *45*(4), 1590–1597. doi:10.1016/j.ejmech.2010.01.002 PMID:20110136

Das, R. N., & Roy, K. (2013). Advances in QSPR/QSTR models of ionic liquids for the design of greener solvents of the future. *Molecular Diversity*, *17*(1), 151–196. doi:10.1007/s11030-012-9413-y PMID:23325356

Dearden, J. C. (Ed.). (1983). *Quantitative approaches to drug design*. Amsterdam, The Netherlands: Elsevier.

Dearden, J. C. (1987). QSAR: end of an era or coming of age? In E. Mutschler & E. Winterfeldt (Eds.), *Trends in medicinal chemistry* (pp. 109–113). Weinheim, Germany: VCH.

Dearden, J. C. (1999). The prediction of melting point. In M. Charton & B. I. Charton (Eds.), *Advances in quantitative structure-property relationships* (Vol. 2, pp. 127–175). Stamford, CT: JAI Press. doi:10.1016/S1874-527X(99)80005-X

Dearden, J. C. (2002). Prediction of environmental toxicity and fate using quantitative structure-activity relationships (QSARs). *Journal of the Brazilian Chemical Society*, *13*(6), 754–762. doi:10.1590/S0103-50532002000600005

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Dearden, J. C. (2003). Quantitative structure-property relationships for prediction of boiling point, vapor pressure, and melting point. *Environmental Toxicology and Chemistry*, *22*(8), 1696–1709. doi:10.1897/01-363 PMID:12924571

Dearden, J. C. (2006). *In silico* prediction of aqueous solubility. *Expert Opinion on Drug Discovery*, *1*(1), 31–52. doi:10.1517/17460441.1.1.31 PMID:23506031

Dearden, J. C. (2010). Expert systems for toxicity prediction. In M. T. D. Cronin & J. C. Madden (Eds.), *In silico toxicology: principles and applications* (pp. 478–507). London, UK: Royal Society of Chemistry. doi:10.1039/9781849732093-00478

Dearden, J. C. (2012). Prediction of physicochemical properties. In B. Reisfeld, & A.N. Mayeno A. (Eds.), Computational toxicology (pp. 93-138). New York, NY: Springer (Humana Press). doi:10.1007/978-1-62703-050-2_6

Dearden, J. C., Cronin, M. T. D., & Kaiser, K. L. E. (2009). How not to develop a quantitative structure-activity or structure-property relationship (QSAR/QSPR). *SAR and QSAR in Environmental Research*, *20*(3-4), 241–266. doi:10.1080/10629360902949567 PMID:19544191

Dearden, J. C., Cronin, M. T. D., & Lappin, D. C. (2007). A comparison of commercial available software for the prediction of pKa. *The Journal of Pharmacy and Pharmacology*, *59*(Suppl. 1), A-7.

Dearden, J. C., & George, E. (1979). Anti-inflammatory potencies of some aspirin derivatives: A quantitative structure-activity study. *The Journal of Pharmacy and Pharmacology*, *31*(S1Suppl. 1), 45P. doi:10.1111/j.2042-7158.1979.tb11593.x PMID:42720

Dearden, J. C., Netzeva, T. I., & Bibby, R. (2003). A comparison of commercially available software for the prediction of partition coefficient. In M. Ford, D. Livingstone, J. Dearden, & H. van de Waterbeemd (Eds.), *Designing drugs and crop protectants: processes, problems and solutions* (pp. 169–170). Oxford, UK: Blackwell.

Dearden, J. C., Rotureau, P., & Fayet, G. (2013). QSPR prediction of physico-chemical properties for REACH. *SAR and QSAR in Environmental Research*, *24*(4), 279–318. doi:10.1080/1062936X.2013.773372 PMID:23521394

Dearden, J. C., & Rowe, P. H. (2015). Use of artificial neural networks in the QSAR prediction of physicochemical properties and toxicities for REACH legislation. In H. Cartwright (Ed.), *Artificial neural networks* (pp. 65–88). New York, NY: Springer. doi:10.1007/978-1-4939-2239-0_5

Dearden, J. C., & Townend, M. S. (1978). Digital computer simulation of the drug transport process. In R. Franke & P Oehme (Eds.), Quantitative structure-activity analysis (pp. 387-393). Berlin, German Democratic Republic: Akademie-Verlag.

Dearden, J. C., & Townend, M. S. (1983). Duration of action of drugs as a function of lipophilicity. In J. C. Dearden (Ed.), *Quantitative approaches to drug design* (pp. 219–220). Amsterdam, The Netherlands: Elsevier, Amsterdam.

Dearden, J. C., & Townend, M. S. (n. d.). Unpublished information.

Devillers, J. (Ed.). (1996). *Genetic algorithms in molecular modeling*. London, UK: Academic Press.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Döbereiner, J. W. (1817). Folgerungen aus seinen stöichimetrischen Versuchen. *Annalen der Physik*, *57*, 435–438.

Döbereiner, J. W. (1829). Versuch zu einer Gruppirung der elementarum Stoffe nach ihrer Analogie. *Annalen der Physik und Chemie*, *91*(2), 301–307. doi:10.1002/andp.18290910217

Dohme, A. R. L., Cox, E. H., & Miller, E. (1926). The preparation of the acyl and alkyl derivatives of resorcinol. *Journal of the American Chemical Society*, *48*(6), 1688–1693. doi:10.1021/ja01417a035

Domine, D., Devillers, J., Chastrette, M., & Karcher, W. (1993). Nonlinear mapping for structure-activity and structure-property modelling. *Journal of Chemometrics*, *7*(4), 227–242. doi:10.1002/cem.1180070402

Doweyko, A. M. (2008). QSAR: Dead or alive? *Journal of Computer-Aided Molecular Design*, *22*(2), 81–89. doi:10.1007/s10822-007-9162-7 PMID:18189157

DRAGON. (n. d.). Retrieved from www.talete.mi.it/products/dragon-descriptors.htm

Du, Q. S., Huang, R. B., Wei, Y. T., Pang, Z. W., Du, L. Q., & Chou, K. C. (2009). Fragment-based quantitative structure-activity relationship (FB-QSAR) for fragment-based drug design. *Journal of Computational Chemistry*, *30*(2), 295–304. doi:10.1002/jcc.21056 PMID:18613071

Dudek, A. Z., Arodz, T., & Gálvez, J. (2006). Computational methods in developing quantitative structure-activity relationships (QSAR): A review. *Combinatorial Chemistry & High Throughput Screening*, *9*(3), 213–228. doi:10.2174/138620706776055539 PMID:16533155

Dujardin-Beaumetz, G.O., & Audigé, I. (1876). De l'action toxique des alcools méthylique, caprylique, aenanthylique et cétylique. *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences*, *83*, 80–81.

Dumas, J. B. A. (1857). Sur les equivalents des corps simples. *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences*, *45*, 709–731.

Egloff, G., Sherman, J., & Dull, R. B. (1940). Boiling point relationships among aliphatic hydrocarbons. *Journal of Physical Chemistry*, *44*(6), 730–745. doi:10.1021/j150402a006

Enoch, S. J., Madden, J. C., & Cronin, M. T. D. (2008). Identification of mechanisms of toxic action for skin sensitisation using a SMARTS pattern based approach. *SAR and QSAR in Environmental Research*, *19*(5-6), 555–578. doi:10.1080/10629360802348985 PMID:18853302

Fauchère, J. L. (Ed.). (1989). *QSAR: quantitative structure-activity relationships in drug design*. New York, NY: Alan R. Liss Inc.

Ferguson, J. (1939). The use of chemical potentials as indices of toxicity. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *127*(848), 387–404. doi:10.1098/rspb.1939.0030

Ferguson, J., & Pirie, H. (1948). The toxicity of vapours to the grain weevil. *Annals of Applied Biology*, *35*(4), 532–550. doi:10.1111/j.1744-7348.1948.tb07396.x

Fieser, L. F., Ettlinger, M. G., & Fawaz, G. (1948). Naphthoquinone antimalarials. XV. Distribution between organic solvents and aqueous buffers. *Journal of the American Chemical Society*, *70*(10), 3228–3232. doi:10.1021/ja01190a008 PMID:18891827

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Fieser, L. F., & Richardson, A. P. (1948). Naphthoquinone antimalarials. II. Correlation of structure and activity against *P. lophurae* in ducks. *Journal of the American Chemical Society*, *70*(10), 3156–3165. doi:10.1021/ja01190a002 PMID:18891813

Ford, M., Livingstone, D., Dearden, J., & van de Waterbeemd, H. (Eds.). (2003). *Designing drugs and crop protectants: processes, problems and solutions*. Oxford, UK: Blackwell.

Franke, R., & Oehme, P. (1973). Current problems of research on active substances. 1. Determination of quantitative structure-activity relationship in biological active substances – theoretical basis, requirements and present status. *Die Pharmazie*, *28*, 489–508. PMID:4131698

Franke, R., & Oehme, P. (Eds.). (1978). Quantitative structure-activity analysis. Berlin, German Democratic Republic: Akademie-Verlag.

Free, S. M., & Wilson, J. W. (1964). A mathematical contribution to structure activity studies. *Journal of Medicinal Chemistry*, *7*(4), 395–399. doi:10.1021/jm00334a001 PMID:14221113

Free, S. M., & Wilson, J. W. (1964). A mathematical contribution to structure-activity studies. *Journal of Medicinal Chemistry*, *7*(4), 395–399. doi:10.1021/jm00334a001 PMID:14221113

Fühner, H. (1903). Über die Einwirkung verschiedener Alkohole auf die Entwicklung der Seeigel. *Archiv für Experimentelle Pathologie und Pharmakologie*, *51*, 1–10.

Fühner, H. (1904). Pharmakologische Studien an Seeigeleiern. Der Wirkungsgrad der Alkohole. *Archiv für Experimentelle Pathologie und Pharmakologie*, *52*, 69–82.

Fujita, T., & Ban, T. (1971). Structure-activity study of phenethylamines as substrates of biosynthetic enzymes of sympathetic transmitters. *Journal of Medicinal Chemistry*, *14*(2), 148–152. doi:10.1021/jm00284a016 PMID:5544401

Furuhama, A., Toida, T., Nishikawa, N., Aoki, Y., Yoshioka, Y., & Shiraishi, H. (2010). Development of an ecotoxicity QSAR model for the KAshinhou Tool for Ecotoxicity (KATE) system, March 2009 version. *SAR and QSAR in Environmental Research*, *21*(5-6), 403–413. doi:10.1080/1062936X.2010.501815 PMID:20818579

Gay-Lussac, J.-L. (1809). Sur la combinaison des substances gazeuses, les unes avec les autres. *Mémoires de la Société de Physique et de Chimie de la Société d'Arcueil*, *2*, 207–234.

George, J. M., & Kier, L. B. (1973). Conformation of thyrotropin-releasing hormone: Prediction from molecular orbital theory. *Journal of Theoretical Biology*, *40*(2), 393–396. doi:10.1016/0022-5193(73)90140-9 PMID:4747250

Giaginis, C., Tsantili-Kakoulidou, A., & Theocharis, S. (2014). Applying quantitative structure-activity relationship (QSAR) methodology for modeling post-mortem redistribution of benzodiazepines and tricyclic antidepressants. *Journal of Analytical Toxicology*, *38*(5), 242–248. doi:10.1093/jat/bku025 PMID:24682110

Godavarthy, S. S., Robinson, R. L., & Gasem, K. A. M. (2006). An improved structure-property model for predicting melting-point temperatures. *Industrial & Engineering Chemistry*, *45*(14), 5117–5126. doi:10.1021/ie051130p

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Golbraikh, A., & Tropsha, A. (2002). Beware of q^2! *Journal of Molecular Graphics & Modelling*, *20*(4), 269–276. doi:10.1016/S1093-3263(01)00123-1 PMID:11858635

González, M. P., Terán, C., Saíz-Urra, L., & Teijeira, M. (2008). Variable selection methods in QSAR: An overview. *Current Topics in Medicinal Chemistry*, *8*(18), 1606–1627. doi:10.2174/156802608786786552 PMID:19075770

Gordon Research Conferences. Available at www.grc.org

Gramatica, P. (2007). Principles of QSAR models validation: Internal or external. *QSAR & Combinatorial Science*, *26*(5), 694–701. doi:10.1002/qsar.200610151

Greener, M. (2007, June). QSAR: prediction beyond the fourth dimension. *Drug Discovery and Development*, 9. Retrieved from www.dddmag.com/articles/2007/09/qsar-prediction-beyond-fourth-dimension

Gubernator, K., & Bohm, H.-J. (Eds.). (1998). *Structure-based ligand design*. Weinheim, Germany: Wiley-VCH. doi:10.1002/9783527612178

Gundertofte, K., & Jørgensen, F. S. (Eds.). (1999). *Molecular modeling and prediction of bioactivity*. New York, NY: Kluwer Academic/Plenum Publishers.

Hadži, D., & Jerman-Blažič, B. (Eds.). (1987). *QSAR in drug design and toxicology*. Amsterdam, The Netherlands: Elsevier.

Hall, L. H., Mohney, B., & Kier, L. B. (1991). The electrotopological state: An atom index for QSAR. *Quantitative Structure-Activity Relationships*, *10*(1), 43–51. doi:10.1002/qsar.19910100108

Hammett, L. P. (1935). Some relations between reaction rates and equilibrium constants. *Chemical Reviews*, *17*(1), 125–136. doi:10.1021/cr60056a010

Hammett, L. P. (1937). The effect of structure upon the reactions of organic compounds. Benzene derivatives. *Journal of the American Chemical Society*, *59*(1), 96–103. doi:10.1021/ja01280a022

Hamzeh-Mivehroud, M., Sokouti, B., & Dastmalchi, S. (2015). An introduction to the basic concepts in QSAR-aided drug design. In K. Roy (Ed.), *Quantitative structure-activity relationships in drug design, predictive toxicology, and risk assessment* (pp. 1–47). Hershey, PA: Medical Information Science Reference. doi:10.4018/978-1-4666-8136-1.ch001

Hansch, C. (1969). A quantitative approach to biochemical structure-activity relationships. *Accounts of Chemical Research*, *2*(8), 232–239. doi:10.1021/ar50020a002

Hansch, C., & Clayton, J. M. (1973). Lipophilic character and biological activity of drugs II. Parabolic case. *Journal of Pharmaceutical Sciences*, *62*(1), 1–21. doi:10.1002/jps.2600620102 PMID:4566530

Hansch, C., & Dunn, W. J. (1972). Linear relationship between lipophilic character and biological activity of drugs. *Journal of Pharmaceutical Sciences*, *61*(1), 1–19. doi:10.1002/jps.2600610102 PMID:4550859

Hansch, C., & Fujita, T. (1964). A method for the correlation of biological activity and chemical structure. *Journal of the American Chemical Society*, *86*(8), 1616–1626. doi:10.1021/ja01062a035

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Hansch, C., & Leo, A. (1979). *Substituent constants for correlation analysis in chemistry and biology*. New York, NY: John Wiley & Sons.

Hansch, C., Maloney, P. P., Fujita, T., & Muir, R. M. (1962). Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients. *Nature*, *194*(4824), 178–180. doi:10.1038/194178b0

Hansch, C., Quinlan, J. E., & Lawrence, G. L. (1968). The linear free-energy relationship between partition coefficients and aqueous solubility of organic liquids. *The Journal of Organic Chemistry*, *33*(1), 347–350. doi:10.1021/jo01265a071

Hansen, O. R., Hallberg, B., Ronquist, G., Briggs, T., Haslewood, G. A. D., & Flood, H. (1962). Hammett series with biological activity. *Acta Chemica Scandinavica*, *16*, 1593–1600. doi:10.3891/acta.chem.scand.16-1593

Hasegawa, K., Kimura, T., & Funatsu, K. (2009). Inverse QSAR study using evolutionary algorithm. *Journal of Computer Aided Chemistry*, *10*, 10–15. doi:10.2751/jcac.10.10

Hermens, J. L. M., & Opperhuizen, A. (Eds.). (1991). *QSAR in environmental toxicology – IV*. New York, NY: Elsevier.

Hewitt, M., Ellison, C. E., Enoch, S. J., Madden, J. C., & Cronin, M. T. D. (2010). Integrating (Q)SAR models, expert systems and read-across approaches for the prediction of developmental toxicity. *Reproductive Toxicology (Elmsford, N.Y.)*, *30*(1), 147–160. doi:10.1016/j.reprotox.2009.12.003 PMID:20006701

Hinrichs, G. D. (1867). *Programm der Atomechanik oder die Chemie eine Mechanik de Pantome*. Iowa City, IA: Augustus Hageboek.

Hinrichs, G. D. (1869). On the classification and the atomic weights of the so-called chemical elements, with reference to Stas' determinations. *Proceedings of the American Association for the Advancement of Science*, *18*, 112–124.

Höltje, H.-D., & Sippl, W. (Eds.), (2001). *Rational approaches to drug design*. Barcelona, Spain: Prous Science.

Holt, J. J. H. (1916). The cockroach: Its destruction and dispersal. *Lancet*, *187*(4840), 1136–1137. doi:10.1016/S0140-6736(01)12843-6

HYBOT. (n. d.). Retrieved from http://molpro.ipac.ru/hybot.html

Ivanciuc, O., Ivanciuc, T., & Cabrol-Bass, D. (2001). Comparative receptor surface analysis (CoRSA) model for calcium channel antagonists. *SAR and QSAR in Environmental Research*, *12*(1-2), 93–111. doi:10.1080/10629360108035373 PMID:11697062

Jaffé, H. H. (1953). A reexamination of the Hammett equation. *Chemical Reviews*, *53*(2), 191–261. doi:10.1021/cr60165a003

Jiang, Y. R., Yang, Y. Y., Chen, Y. L., & Liang, Z. J. (2013). CoMFA, CoMSIA and HQSAR studies of acetylcholinesterase inhibitors. *Current Computer-aided Drug Design*, *9*(3), 385–395. doi:10.2174/15734099113099990015 PMID:24010934

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

John, S., Thangapandian, S., Arooj, M., Hong, J. C., Kim, K. D., & Lee, K. W. (2011). Development, evaluation and application of 3D QSAR pharmacophore model in the discovery of potential human renin inhibitors. *BMC Bioinformatics*, *12*(Supp. 14), S4. www.biomedcentral.com/1471-2105/12/S14/S4 doi:10.1186/1471-2105-12-S14-S4 PMID:22372967

Kaiser, K. L. E. (Ed.). (1984). *QSAR in environmental toxicology*. Dordrecht, The Netherlands: D. Reidel. doi:10.1007/978-94-009-6415-0

Kaiser, K. L. E. (Ed.). (1987). *QSAR in environmental toxicology – II*. Dordrecht, The Netherlands: D. Reidel. doi:10.1007/978-94-009-3937-0

Kamlet, M. J., Abboud, J.-L. M., Abraham, M. H., & Taft, R. W. (1983). Linear solvation energy relationships. 23. A comprehensive collection of the solvatochromic parameters, π*, α, and β, and some methods for simplifying the generalized solvatochromic equation. *The Journal of Organic Chemistry*, *48*(17), 2877–2887. doi:10.1021/jo00165a018

Kar, S., & Roy, K. (2012). QSAR of phytochemicals for the design of better drugs. *Expert Opinion on Drug Discovery*, *7*(10), 877–902. doi:10.1517/17460441.2012.716420 PMID:22897485

Kauffman, G. H. (1999). From triads to catalysis: Johann Wolfgang Döbereiner (1780-1849) on the 150th anniversary of his death. *Chemical Educator*, *4*(5), 186–197. doi:10.1007/s00897990326a

Kier, L. B. (1967). Molecular orbital calculation of the preferred conformation of acetylcholine, muscarine and muscarone. *Molecular Pharmacology*, *3*, 487–494. PMID:6052710

Kier, L. B., & Hall, L. H. (1976). *Molecular connectivity in chemistry and drug research*. New York, NY: Academic Press.

Kier, L. B., & Hall, L. H. (1986). *Molecular connectivity in structure activity analysis*. London, UK: John Wiley.

Kim, K. H., Greco, G., Novellino, E., Silipo, C., & Vittoria, A. (1993). Use of the hydrogen bond potential function in a comparative molecular field analysis (CoMFA) on a set of benzodiazepines. *Journal of Computer-Aided Molecular Design*, *7*(3), 263–280. doi:10.1007/BF00125502 PMID:8377024

Klarmann, E., Gatyas, L. W., & Shternov, V. A. (1931). Bactericidal properties of monoethers of dihydric phenols. I. The monoethers of resorcinol. *Journal of the American Chemical Society*, *53*(9), 3397–3407. doi:10.1021/ja01360a025

Klebe, G., & Abraham, U. (1999). Comparative molecular similarity index analysis (CoMSIA) to study hydrogen-bonding properties and to score combinatorial libraries. *Journal of Computer-Aided Molecular Design*, *13*(1), 1–10. doi:10.1023/A:1008047919606 PMID:10087495

Knoll, J., & Darvas, F. (Eds.). (1980). *Chemical structure-biological activity relationships: quantitative approaches*. Budapest, Hungary: Akadémiai Kiadó.

Könemann, H. (1981). Quantitative structure-activity relationships in fish toxicity studies. Part 1: Relationship for 50 industrial pollutants. *Toxicology*, *19*(3), 209–221. doi:10.1016/0300-483X(81)90130-X PMID:7233445

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Kopp, H. (1842). On a great regularity in the physical properties of analogous organic compounds. *Philosophical Magazine (Series 3)*, *20*(130), 187–197. doi:10.1080/14786444208650552

Kopp, H. (1844). Ueber den Zusammenhang zwischen der chemischen Constitution und einigen physikalischen Eigenschaften bei flüssigen Verbingungen. *Justus Liebigs Annalen der Chemie*, *50*(1), 71–144. doi:10.1002/jlac.18440500105

Kroemer, R. T. (2007). Structure-based drug design: Docking and scoring. *Current Protein & Peptide Science*, *8*, 312–328. doi:10.2174/138920307781369382 PMID:17696866

Kubinyi, H. (1977). Quantitative structure-activity relationships. 7. The bilinear model, a new model for nonlinear dependence of biological activity on hydrophobic character. *Journal of Medicinal Chemistry*, *20*(5), 625–629. doi:10.1021/jm00215a002 PMID:857018

Kubinyi, H. (1993). *QSAR: Hansch analysis and related approaches*. Weinheim, Germany: VCH. doi:10.1002/9783527616824

Kubinyi, H. (2002). From narcosis to hyperspace: The history of QSAR. *Quantitative Structure-Activity Relationships*, *21*(4), 348–356. doi:10.1002/1521-3838(200210)21:4<348::AID-QSAR348>3.0.CO;2-D

Leach, A. G., & Gillet, V. J. (2007). *An introduction to chemoinformatics*. Dordrecht, The Netherlands: Springer. doi:10.1007/978-1-4020-6291-9

Leach, A. G., Jones, H. D., Cosgrove, D. A., Kenny, P. W., Ruston, L., MacFaul, P., & Law, B. et al. (2006). Matched molecular pairs as a guide in the optimization of pharmaceutical properties: A study of aqueous solubility, plasma protein binding and oral exposure. *Journal of Medicinal Chemistry*, *49*(23), 6672–6682. doi:10.1021/jm0605233 PMID:17154498

Leahy, D. E., Morris, J. J., Taylor, P. J., & Wait, A. R. (1992). Model solvent systems for QSAR. 3. An LSER analysis of the critical quartet – new light on hydrogen-bond strength and directionality. *Journal of the Chemical Society – Perkins Transactions*, *2*(4), 705–722. doi:10.1039/p29920000705

Leo, A., Hansch, C., & Church, C. (1969). Comparison of parameters currently used in the study of structure-activity relationships. *Journal of Medicinal Chemistry*, *12*(5), 766–771. doi:10.1021/jm00305a010 PMID:5812184

Leo, A., Jow, P. Y. C., Silipo, C., & Hansch, C. (1975). Calculation of hydrophobic constant (log P) from π and *f* constants. *Journal of Medicinal Chemistry*, *18*(9), 865–868. doi:10.1021/jm00243a001 PMID:1159707

Li, L., Hu, J., & Ho, Y.-S. (2014). Global performance and trend of QSAR/QSPR research: A bibliometric analysis. *Molecular Informatics*, *33*(10), 655–668. doi:10.1002/minf.201300180

Liang, Y.-Z., Kvalheim, O. M., & Manne, R. (1993). White, Grey and black multicomponent systems: A classification of mixture problems and methods for their quantitative analysis. *Chemometrics and Intelligent Laboratory Systems*, *18*(3), 235–250. doi:10.1016/0169-7439(93)85001-W

Liao, C., & Nicklaus, M. C. (2009). Comparison of nine programs predicting pKa values of pharmaceutical substances. *Journal of Chemical Information and Modeling*, *49*(12), 2801–2812. doi:10.1021/ci900289x PMID:19961204

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Lipnick, R. L. (1990). Narcosis: fundamental and baseline toxicity mechanism for nonelectrolyte organic chemicals. In W. Karcher & J. Devillers (Eds.), *Practical applications of quantitative structure-activity relationships (QSAR) in environmental chemistry and toxicology,* Dordrecht, The Netherlands (pp. 281–293). Kluwer Academic Publishers.

Lipnick, R. L. (Ed.). (1991). *Studies of narcosis: Charles Ernest Overton*. London, UK: Chapman & Hall. doi:10.1007/978-94-011-3096-7

Lithner, G. (1989). Some fundamental relationships between metal toxicity in freshwater, physico-chemical properties and background levels. *The Science of the Total Environment*, *87-88*, 356–380. doi:10.1016/0048-9697(89)90249-0

Livingstone, D. (1995). *Data analysis for chemists*. Oxford, UK: Oxford University Press.

Livingstone, D. J. (2004). Building QSAR models: a practical guide. In M. T. D. Cronin & D. J. Livingstone (Eds.), *Predicting chemical toxicity and fate* (pp. 151–170). Boca Raton, FL: CRC Press. doi:10.1201/9780203642627.ch7

Livingstone, D. J., Ford, M. G., Huuskonen, J. J., & Salt, D. W. (2001). Simultaneous prediction of aqueous solubility and octanol/water partition coefficient based on descriptors derived from molecular structure. *Journal of Computer-Aided Molecular Design*, *15*(8), 741–752. doi:10.1023/A:1012284411691 PMID:11718478

Luis, P., Garea, A., & Irabien, A. (2010). Quantitative structure-activity relationships (QSARs) to estimate ionic liquids ecotoxicity EC_{50} (*Vibrio fischeri*). *Journal of Molecular Liquids*, *152*(1-3), 28–33. doi:10.1016/j.molliq.2009.12.008

Mannhold, R., & Dross, K. (1996). Calculation procedures for molecular lipophilicity: A comparative study. *Quantitative Structure-Activity Relationships*, *15*(5), 403–409. doi:10.1002/qsar.19960150506

Martin, Y., & Stouch, T. (2011). In tribute to Corwin Hansch, father of QSAR. *Journal of Computer-Aided Molecular Design*, *25*(6), 491. doi:10.1007/s10822-011-9449-6 PMID:21710390

Martin, Y. C. (1978). *Quantitative drug design: a critical introduction*. New York, NY: Marcel Dekker.

Martin, Y. C. (2002). 3D QSAR: current state, scope, and limitations. In H. Kubinyi, G. Folkers, & Y.C. Martin (Eds.), 3D QSAR in drug design: recent advances (pp. 3-23). New York, NY: Kluwer Academic Publishers.

Martin, Y. C. (2010). *Quantitative drug design: a critical introduction* (2nd ed.). Boca Raton, FL: CRC Press. doi:10.1201/9781420071009

Martin, Y. C. (2012). Hansch analysis 50 years on. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, *2*, 435–442.

Martins, J. P. A., Barbosa, E. G., Pasqualoto, K. F. M., & Ferreira, M. M. C. (2009). LQTA-QSAR:a new 4D-QSAR methodology. *Journal of Chemical Information and Modeling*, *49*(6), 1428–1436. doi:10.1021/ci900014f PMID:19422246

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Matthews, E. J., Kruhlak, N. L., Benz, R. D., Contrera, J. F., Marchant, C. A., & Yang, C. (2008). Combined use of MC4PC, MDL-QSAR, BioEpisteme, Leadscope PDM, and Derek for Windows software to achieve high-performance, high-confidence, mode of action-based predictions of chemical carcinogenesis in rodents. *Toxicology Mechanisms and Methods*, *18*(2-3), 189–206. doi:10.1080/15376510701857379 PMID:20020914

McFarland, J. W. (1970). On the parabolic relationship between drug potency and hydrophobicity. *Journal of Medicinal Chemistry*, *13*(6), 1192–1196. doi:10.1021/jm00300a040 PMID:5479865

McGowan, J. C. (1952). The physical toxicity of chemicals. II. Factors affecting physical toxicity in aqueous solutions. *Journal of Chemical Technology and Biotechnology (Oxford, Oxfordshire)*, *2*, 323–328.

McGowan, J. C., & Mellors, A. (1986). *Molecular volumes in chemistry and biology*. Chichester, UK: Ellis Horwood.

Mekenyan, O., Nikolova, N., Karabunarliev, S., Bradbury, S. P., Ankley, G. T., & Hansen, B. (1999). New developments in a hazard identification algorithm for hormone receptor ligands. *Quantitative Structure-Activity Relationships*, *18*(2), 139–153. doi:10.1002/(SICI)1521-3838(199906)18:2<139::AID-QSAR139>3.0.CO;2-K

Mendeleev, D. (1869). The relation between the properties and atomic weights of the elements.[English translation]. *Journal of the Russian Chemical Society*, *1*, 60–77.

Meng, X. Y., Zhang, H. X., Mezei, M., & Cui, M. (2011). Molecular docking: A powerful approach for structure-based drug design. *Current Computer-aided Drug Design*, *7*(2), 146–157. doi:10.2174/157340911795677602 PMID:21534921

Meyer, H. (1899). Zur Theorie der Alkoholnarkose. Erste Mitteilung. Welche Eigenschaft der Anasthetica bedingt ihre narkotische Wirkung? *Naunyn-Schmiedebergs Archiv fur Experimentelle Pathologie und Pharmakologie*, *42*(2-4), 109–118. doi:10.1007/BF01834479

Millay, E. St. V. (1988). *Collected sonnets*. New York: Harper Perennial.

Mills, E. J. (1884a). On melting point and boiling point as related to composition. *Philosophical Magazine, Series 5*, *17*(105), 173–187. doi:10.1080/14786448408627502

Mills, E. J. (1884b). On the numerics of the elements. *Philosophical Magazine, Series 5*, *18*(114), 393–399. doi:10.1080/14786448408627612

Mitra, I., Saha, A., & Roy, K. (2012). Development of multiple QSAR models for consensus predictions and unified mechanistic interpretations of the free-radical scavenging activities of chromone derivatives. *Journal of Molecular Modeling*, *18*(5), 1819–1840. doi:10.1007/s00894-011-1198-x PMID:21850572

Modarresi, H., Modarress, H., & Dearden, J. C. (2007). QSPR model of Henry's law constant for a diverse set of organic chemicals based on genetic algorithm-radial basis function network approach. *Chemosphere*, *66*(11), 2067–2076. doi:10.1016/j.chemosphere.2006.09.049 PMID:17113627

MOE. (n. d.). Retrieved from www.chemcomp.com

Molconn-Z. (n. d.) Retrieved from www.edusoft.lc.com/molconn/

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Moore, W. (1917a). The toxicity of various benzene derivatives to insects. *Journal of Agricultural Research*, *9*, 371–381.

Moore, W. (1917b). Volatility of organic compounds as an index of the toxicity of their vapors to insects. *Journal of Agricultural Research*, *10*, 365–371.

Moore, W. (1918). Physical properties governing the efficacy of contact insecticides. *Journal of Agricultural Research*, *13*, 523–538.

Mostrag-Szlichtyng, A., Zaldivar-Comenges, J.-M., & Worth, A. P. (2010). Computational toxicology at the European Commission's Joint Research Centre. *Expert Opinion on Drug Metabolism & Toxicology*, *6*(7), 785–792. doi:10.1517/17425255.2010.489551 PMID:20443752

Muratov, E. N., Varlamova, E. K., Artemenko, A. G., Polishchuk, P. G., & Kuz'min, V. E. (2012). Existing and developing approaches for QSAR analysis of mixtures. *Molecular Informatics*, *31*(3-4), 202–221. doi:10.1002/minf.201100129

Muratov, E. N., Varlamova, E. V., Kuzmin, V. E., Artemenko, A. G., Muratov, N. N., Mileyko, S., & Tropsha, A. et al. (2014). "Everything out" validation approach for Qsar models of chemical mixtures. *Journal of Clinical Pharmaceutics*, *1*(1), 1005. Retrieved from www.jscimedcentral.com/ClinicalPharmaceutics/clinicalpharmaceutics-1-1005.pdf

Neely, W. B., Branson, D. R., & Blau, G. E. (1981). Partition coefficient to measure bioconcentration potential of organic chemicals in fish. *Environmental Science & Technology*, *8*(13), 1113–1115. doi:10.1021/es60098a008

Netzeva, T. I., Worth, A., Aldenberg, T., Benigni, R., Cronin, M. T., Gramatica, P., & Yang, C. et al. (2005).Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships. The report and recommendations of ECVAM Workshop 52. *Alternatives to Laboratory Animals*, *33*, 155–173. PMID:16180989

Newlands, J. A. R. (1864). Relations among equivalents. *Chemical News*, *10*, 59–60.

Nys, G. G., & Rekker, R. F. (1973). Statistical analysis of a series of partition coefficients with special reference to predictability of folding of drug molecules – introduction of hydrophobic fragmental constants. *Chimica Therapeutica*, *8*, 521–535.

Odling, W. (1857). On the natural groupings of the elements. *Philosophical Magazine*, Series 4, *13*, 423-439; 480-497.

OECD. (2007). *Guidance document on the validation of (quantitative) structure-activity relationships: (Q)SAR models.* Retrieved from www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/jm/mono(2007)2

OECD QSAR Toolbox. Available at www.qsartoolbox.org

Oksel, C., Ma, C. Y., Liu, J. J., Wilkins, T., & Wang, X. Z. (2015). (Q)SAR modelling of nanomaterial toxicity: A critical review. *Particuology*, *21*, 1–19. doi:10.1016/j.partic.2014.12.001

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Overton, C. E. (1897). Osmotic properties of cells in the bearing on toxicology and pharmacology. *Zeitschrift für Physikalische Chemie*, *22*, 189–209.

Overton, C. E. (1899). Ueber die allgemeinen osmotischen Eigenschaften der Zelle, ihre vermutlichen Ursachen und ihre Bedeutung für die Physiologie. *Vierteljahrschrift der Naturforschenden Gesellschaft in Zürich*, *44*, 88–135.

Overton, C. E. (1901). *Studien über die Narkose*. Jena: Gustav Fischer.

Parascandola, J. Structure-activity relationships – the early mirage. (1971). *Pharmacy in History*, *13*, 3-10. PMID:11609444

Penniston, J. T., Beckett, L., Bentley, D. L., & Hansch, C. (1969). Passive permeation of organic compounds through biological tissues: A non-steady state theory. *Molecular Pharmacology*, *5*, 333–341. PMID:5803385

Peristera, O., Spreafico, M., Smiesko, M., Ernst, B., & Vedani, A. (2009). Mixed-model QSAR at the human mineralocorticoid receptor: Predicting binding mode and affinity of anabolic steroids. *Toxicology Letters*, *189*(3), 219–224. doi:10.1016/j.toxlet.2009.05.025 PMID:19523507

Perkins, R., Fang, H., Tong, W., & Welsh, W. J. (2003). QSAR methods: Perspectives on drug discovery and toxicology. *Environmental Toxicology and Chemistry*, *22*, 1666–1679. doi:10.1897/01-171 PMID:12924569

Polanski, J. (2009). Receptor-dependent multidimensional QSAR for modeling drug-receptor interactions. *Current Medicinal Chemistry*, *16*(25), 3243–3257. doi:10.2174/092986709788803286 PMID:19548875

Polanski, J., Gasteiger, J., & Jarzembek, K. (2000). Self-organizing neural networks for screening and development of novel artificial sweetener candidates. *Combinatorial Chemistry & High Throughput Screening*, *3*(6), 481–495. doi:10.2174/1386207003331427 PMID:11121517

Polishchuk, P. G., Muratov, E. N., Artemenko, A. G., Kolumbin, O. G., Muratov, N. N., & Kuz'min, V. E. (2009). Application of random forest approach to QSAR prediction of aquatic toxicity. *Journal of Chemical Information and Modeling*, *49*(11), 2481–2488. doi:10.1021/ci900203n PMID:19860412

Ponte, J. P. (1992). The history of the concept of function and some mathematical implications. *The Mathematics Educator*, *3*(2), 3–8.

Purcell, W. P., Bass, G. E., & Clayton, J. M. (1973). *Strategy of drug design: a guide to biological activity*. New York, NY: John Wiley & Sons.

Puzyn, T., Gajewicz, A., Leszczynska, D., & Leszczynski, J. (2010). Nanomaterials – the next great challenge for QSAR modelers. In T. Puzyn, J. Leszczynski, & M. T. D. Cronin (Eds.), *Recent advances in QSAR studies: methods and applications* (pp. 383–409). Dordrecht, The Netherlands: Springer. doi:10.1007/978-1-4020-9783-6_14

Puzyn, T., Rasulev, B., Gajewicz, A., Hu, X., Dasari, T. P., Michalkova, A., & Leszczynski, J. et al. (2011). Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles. *Nature Nanotechnology*, *6*(3), 175–178. doi:10.1038/nnano.2011.10 PMID:21317892

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Rabuteau, A. (1870). De quelques propriétés nouvelles ou peu connues de l'alcool du vin ou l'alcool ethylique; déductions thérapeutiques de ces propriétés - des effets toxiques des alcools butylique et amylique – application à l'alcoolisation du vin improprement appelée vinage. *L'Union Medicale (3rd Series)*, 10, 165-173.

Raevsky, O. A. (1984). *An introduction to computer-aided design of biologically active compounds*. Moscow, Russia: Moscow Chimico-Technology Institute. (in Russian)

Raevsky, O. A., Grigor'ev, V. Yu., Polianczyk, D. E., Raevskaja, O. E., & Dearden, J. C. (2014). Calculation of aqueous solubility of crystalline unionized organic chemicals and drugs based on structural similarity and physicochemical descriptors. *Journal of Chemical Information and Modeling*, *54*(2), 683–691. doi:10.1021/ci400692n PMID:24456022

REACH legislation. (2007). Retreived from http://echa.europa.eu/regulations/reach/legislation

Richardson, B. W. (1869). Physiological research on alcohols. *Medical Times and Gazette*, *2*, 703–706.

Richarz, A. N., Berthold, M. R., Fioravanzo, E., Neagu, D., Yang, C., Zaldivar-Comenges, J.-M., & Cronin, M. T. D. (2013). Integrated *in silico* modelling for the prediction of chronic toxicity. *International Journal of Toxicology*, *32*, 72–73.

Richarz, A.-N., Madden, J. C., Marchese Robinson, R. L., Lubiński, L., Mokshina, E., Urbaszek, P., & Cronin, M. T. D. et al. (2015). Development of computational methods for the prediction of toxicity of nanomaterials. *Perspectives on Science*, *3*, 27–29. doi:10.1016/j.pisc.2014.11.015

Richet, C. (1893). Sur le rapport entre la toxicité et les propriétés physiques des corps. *Comptes Rendus des Séances de la Société de Biologie (Paris)*, *45*, 775–776.

Rogers-Evans, M., Alanine, A. I., Bleicher, K. H., Kube, D., & Schneider, G. (2004). Identification of novel cannabinoid receptor ligands *via* evolutionary do novo design and rapid parallel synthesis. *Molecular Informatics*, *23*, 426–430.

Rowberg, K. A., Even, M., Martin, E., & Hopfinger, A. J. (1994). QSAR and molecular shape analysis of three series of 1-(phenylcarbamoyl)-2-pyrazoline insecticides. *Journal of Agricultural and Food Chemistry*, *42*(2), 374–380. doi:10.1021/jf00038a027

Roy, K., & Ghosh, G. (2004). QSTR with extended topochemical atom indices. 4. Modeling of the acute toxicity of phenylsulfonyl carboxylates to *Vibrio fischeri* using principal component factor analysis and principal component regression analysis. *QSAR & Combinatorial Science*, *23*(7), 526–535. doi:10.1002/qsar.200430891

Roy, K., Kar, S., & Ambure, P. (2015). On a simple approach for determining applicability domain. *Chemometrics and Intelligent Laboratory Systems*, *145*, 22–29. doi:10.1016/j.chemolab.2015.04.013

Roy, K., Kar, S., & Das, R. N. (2015a). *A primer on QSAR/QSPR modeling. Fundamental concepts*. Heidelberg, Germany: Springer.

Roy, K., Kar, S., & Das, R. N. (2015b). *Understanding the basics of QSAR for applications in pharmaceutical sciences and risk assessment*. London, UK: Academic Press.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Sanz, F., Giraldo, J., & Manaut, F. (Eds.). (1985). *QSAR and molecular modelling: concepts, computational tools and biological applications*. Barcelona, Spain: J.R. Prous.

Schaper, K.-J., & Saxena, A. K. (1991). QSAR analysis of dose and property dependent activity classes of homologous drugs. In C. Silipo & A. Vittoria (Eds.), *QSAR: rational approaches to the design of bioactive compounds* (pp. 45–48). Amsterdam, The Netherlands: Elsevier.

Schneider, G., & Fechner, U. (2005). Computer-based *de novo* design of drug-like molecules. *Nature Reviews. Drug Discovery*, *4*(8), 649–663. doi:10.1038/nrd1799 PMID:16056391

Schön, U., Antel, J., Brückner, R., Messinger, J., Franke, R., & Gruska, A. (1998). Synthesis, pharmacological characterization, and quantitative structure-activity relationship analyses of 3,7,9,9-tetraalkylbispidines: Derivatives with specific bradycardic activity. *Journal of Medicinal Chemistry*, *41*(3), 318–331. doi:10.1021/jm970120q PMID:9464363

Schultz, T. W., Lin, D. T., & Arnold, L. M. (1991). QSARs for monosubstituted anilines eliciting the polar narcosis mechanism of action. *The Science of the Total Environment*, *109-110*, 569–580. doi:10.1016/0048-9697(91)90210-6 PMID:1815375

Selassie, C. D. (2003). History of quantitative structure-activity relationships. In D. J. Abraham (Ed.), *Burger's Medicinal Chemistry and Drug Discovery* (6th ed., Vol. 1, pp. 1–48). New York, NY: John Wiley & Sons. doi:10.1002/0471266949.bmc001

Senese, C. L., & Hopfinger, A. J. (2003). Receptor-independent 4D-QSAR analysis of norstatine derived inhibitors of HIV-1 protease. *Journal of Chemical Information and Computer Sciences*, *43*(4), 1297–1307. doi:10.1021/ci0340456 PMID:12870923

Seydel, J. K. (Ed.). (1985). *QSAR and strategies in the design of bioactive compounds*. Weinheim, Federal Republic of Germany: VCH.

Silipo, C., & Vittoria, A. (Eds.). (1991). *QSAR: rational approaches to the design of bioactive compounds*. Amsterdam, The Netherlands: Elsevier.

Singh, K. P., & Gupta, S. (2014). Nano-QSAR modeling for predicting biological activity of diverse nanomaterials. *RSC Advances*, *4*(26), 13215–13230. doi:10.1039/c4ra01274g

Sushko, Y., Novotarskyi, S., Körner, R., Vogt, J., Abdelaziz, A., & Tetko, I. V. (2014). Prediction-driven matched molecular pairs to interpret QSARs and aid the molecular optimization process. *Journal of Cheminformatics*, *6*(1), 48. www.jcheminf.com/contents/6/1/48 doi:10.1186/s13321-014-0048-0 PMID:25544551

Taft, R. W. Jr. (1952). Polar and steric substituent constants for aliphatic and *o*-benzoate groups from rates of esterification and hydrolysis of esters. *Journal of the American Chemical Society*, *74*(12), 3120–3128. doi:10.1021/ja01132a049

Tan, S.-M., Jiao, J., Zhu, X.-L., Zhou, Y.-P., Song, D.-D., Gong, H., & Yu, R.-Q. (2010). QSAR studies of a diverse series of antimicrobial agents against *Candida albicans* by classification and regression trees. *Chemometrics and Intelligent Laboratory Systems*, *103*(2), 184–190. doi:10.1016/j.chemolab.2010.07.005

Thomson, W. (1884). *The practical applications of electricity*. London, U.K.: I.C.E.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Tichý, M. (Ed.). (1976). *Chemical structure-biological activity relationships*. Budapest, Hungary: Akadémiai Kiadó. doi:10.1007/978-3-0348-5795-6

Tichý, M. (Ed.). (1985). *QSAR in toxicology and xenobiochemistry*. Amsterdam, The Netherlands: Elsevier.

Tilley, F. W., & Schaffer, J. M. (1926). Relation between the chemical constitution and germicidal activity of the monohydric alcohols and phenols. *Journal of Bacteriology*, *12*, 303–309. PMID:16559220

Tilley, F. W., & Schaffer, J. M. (1928). Chemical constitution and germicidal activity of amines, ketones, and aldehydes. *Journal of Bacteriology*, *16*, 279–285. PMID:16559340

Topliss, J. G., & Costello, R. J. (1972). Chance correlations in structure-activity studies using multiple regression analysis. *Journal of Medicinal Chemistry*, *15*(10), 1066–1068. doi:10.1021/jm00280a017 PMID:5069775

Topliss, J. G., & Edwards, R. P. (1979). Chance factors in studies of quantitative structure-activity relationships. *Journal of Medicinal Chemistry*, *22*(10), 1238–1244. doi:10.1021/jm00196a017 PMID:513071

Traube, J. (1904). Theorie der Osmose und Narkose. *(Pfluger's) Archiv für die gesammte*. *Physiologie*, *105*, 541–558.

Tsakovska, I., Lessigiarska, I., Netzeva, T., & Worth, A. P. (2008). A mini review of mammalian (Q) SAR models. *QSAR & Combinatorial Science*, *27*(1), 41–48. doi:10.1002/qsar.200710107

Tuccinardi, T., Ortore, G., Rossello, A., Supuran, C. T., & Martinelli, A. (2007). Homology modeling and receptor-based 3D-QSAR study of carbonic anhydrase IX. *Journal of Chemical Information and Modeling*, *47*(6), 2253–2262. doi:10.1021/ci700214j PMID:17949069

Turner, J. E., England, M. W., Schultz, T. W., & Kwaak, N. J. (Eds.). (1988). QSAR 88. (CONF-880520—(DE88013180)). Oak Ridge, TN: U.S. Dept. of Energy.

Unger, S. H. (1984). Whither QSAR? Wither QSAR? In M. Kuchař (Ed.), *QSAR in design of bioactive compounds* (pp. 1–9). Barcelona, Spain: Prous Publishers.

Valerio, L. G., Kruhlak, N. L., & Benz, R. D. (2010). In silico toxicology screening of the rodent carcinogenic potential of phytochemicals using quantitative structure-activity analysis. In Deutsche Forschungsgemeinschaft, & G. Eisenbrand (Eds.), Risk assessment of phytochemicals in food: novel applications. Weinheim, Germany: Wiley-VCH. doi doi:10.1002/9783527634705.ch3a

van de Waterbeemd, H. (1992). The history of drug research: From Hansch to the present. *Quantitative Structure-Activity Relationships*, *11*(2), 200–204. doi:10.1002/qsar.19920110215

van de Waterbeemd, H., Testa, B., & Folkers, G. (1997). *Computer-assisted lead finding and optimization*. Weinheim, Germany: Wiley-VCH. doi:10.1002/9783906390406

Van Leeuwen, K., Schultz, T. W., Henry, T., Diderich, R., & Veith, G. D. (2009). Using chemical categories to fill data gaps in hazard assessment. *SAR and QSAR in Environmental Research*, *20*(3-4), 207–220. doi:10.1080/10629360902949179 PMID:19544189

van Spronsen, J. W. (1959). The prehistory of the periodic system of the elements. *Journal of Chemical Education*, *36*(11), 565–567. doi:10.1021/ed036p565

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Vedani, A., & Dobler, M. (2002). 5D-QSAR: The key for simulating induced fit? *Journal of Medicinal Chemistry*, *45*(11), 2139–2149. doi:10.1021/jm011005p PMID:12014952

Verma, J., Khedkar, V. M., & Coutinho, E. C. (2010). 3D QSAR in drug design – a review. *Current Topics in Medicinal Chemistry*, *10*(1), 95–115. doi:10.2174/156802610790232260 PMID:19929826

Vernon, H. M. (1911). The action of homologous alcohols and aldehydes on the tortoise heart. *The Journal of Physiology*, *43*(3-4), 325–342. doi:10.1113/jphysiol.1911.sp001475 PMID:16993113

Vyas, V. K., Ukawala, R. D., Ghate, M., & Chintha, C. (2012). Homology modeling a fast tool for drug discovery: Current perspectives. *Indian Journal of Pharmaceutical Sciences*, *74*(1), 1–17. doi:10.4103/0250-474X.102537 PMID:23204616

Walker, J. D., Dearden, J. C., Schultz, T. W., Jaworska, J., & Comber, M. H. I. (2003). QSARs for new practitioners. In J. D. Walker (Ed.), *QSARs for pollution prevention, toxicity screening, risk assessment, and web applications* (pp. 3–18). Pensacola, FL: SETAC Press.

Walker, J. D., Enache, E., & Dearden, J. C. (2007). Quantitative cationic activity relationships for predicting toxicity of metal ions from physicochemical properties and natural occurrence levels. *QSAR & Combinatorial Science*, *26*(4), 522–527. doi:10.1002/qsar.200630016

Walker, J. D., Enache, M., & Dearden, J. C. (2003). Quantitative cationic-activity relationships for predicting toxicity of metals. *Environmental Toxicology and Chemistry*, *22*(8), 1916–1935. doi:10.1897/02-568 PMID:12924590

Walker, J. D., Jaworska, J., Comber, M. H. I., Schultz, T. W., & Dearden, J. C. (2003). Guidelines for developing and using quantitative structure-activity relationships. *Environmental Toxicology and Chemistry*, *22*(8), 1653–1665. doi:10.1897/01-627 PMID:12924568

Warburg, O. (1921). Physikalische Chemie der Zellatmung. *Biochemische Zeitschrift*, *119*, 134–166.

Web of Science. (n. d.). Retrieved from http://wok.com

Wei, J. (1975). Least squares fitting of an elephant. *Chemtech*, *5*, 128–129.

Wells, P. R. (1968). *Linear free energy relationships*. London, UK: Academic Press.

Wermuth, C. G. (Ed.). (1993). *Trends in QSAR and molecular modelling*. Leiden, The Netherlands: ESCOM. doi:10.1007/978-94-011-1472-1

Wikipedia. (n. d.). Dmitri Mendeleev. Retrieved from https://en.wikipedia.org/wiki/Dmitri_Mendeleev

Wold, S., Ruhe, A., Wold, H., & Dunn, W. J. III. (1984). The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. *SIAM Journal on Scientific and Statistical Computing*, *5*(3), 735–743. doi:10.1137/0905052

Wong, W. W. L., & Burkowski, F. J. (2009). A constructive approach for discovering new drug leads: using a kernel methodology for the inverse-QSAR problem. *Journal of Cheminformatics, 2009(1),* 4. Retrieved from www.jcheminf.com/content/1/1/4

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Worth, A. P., & Cronin, M. T. D. (2003). The use of discriminant analysis, logistic regression and classification tree analysis in the development of classification models for human health effects. *Journal of Molecular Structure THEOCHEM*, *622*(1-2), 97–111. doi:10.1016/S0166-1280(02)00622-X

Yaffe, D., Cohen, Y., Espinosa, G., Arenas, A., & Giralt, F. (2001). A fuzzy ARTMAP based on quantitative structure-property relationships (QSPRs) for predicting aqueous solubility of organic compounds. *Journal of Chemical Information and Computer Sciences*, *41*(5), 1177–1207. doi:10.1021/ci010323u PMID:11604019

Yu, R., Wang, J., Wang, R., Lin, Y., Hu, Y., Wang, Y., & Lin, Z. et al. (2015). Combined pharmacophore modeling and docking studies on CYP11B1 inhibitors. *Molecules (Basel, Switzerland)*, *20*, 1014–1030. doi:10.3390/molecules20011014 PMID:25584832

Yuan, H., & Parrill, A. (2005). Cluster analysis and three-dimensional QSAR studies of HIV-1 integrase inhibitors. *Journal of Molecular Graphics & Modelling*, *23*(4), 317–328. doi:10.1016/j.jmgm.2004.10.003 PMID:15670952

Zahradnik, R., & Chvapil, M. (1960). Study of the relationship between the magnitude of biological activity and the structure of aliphatic compounds. *Experientia*, *16*(11), 511–512. doi:10.1007/BF02158373 PMID:13787667

Zapffe, C. A. (1969). Gustavus Hinrichs, precursor of Mendeleev. *Isis*, *60*(4), 461–476. doi:10.1086/350537

This work was previously published in the International Journal of Quantitative Structure-Property Relationships (IJQSPR), 1(1); edited by Kunal Roy, pages 1-44, copyright year 2016 by IGI Publishing (an imprint of IGI Global).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 4
Ligand- and Structure-Based Drug Design of Non-Steroidal Aromatase Inhibitors (NSAIs) in Breast Cancer

Tarun Jha
Jadavpur University, India

Amit Kumar Halder
Jadavpur University, India

Nilanajn Adhikari
Jadavpur University, India

Achintya Saha
University of Calcutta, India

ABSTRACT

Aromatase is a multienzyme complex overexpressed in breast cancer and responsible for estrogen production. It is the potential target for designing anti-breast cancer drugs. Ligand and Structure-Based Drug Designing approaches (LBDD and SBDD) are involved in development of active and more specific Nonsteroidal Aromatase Inhibitors (NSAIs). Different LBDD and SBDD approaches are presented here to understand their utility in designing novel NSAIs. It is observed that molecules should possess a five or six membered heterocyclic nitrogen containing ring to coordinate with heme portion of aromatase for inhibition. Moreover, one or two hydrogen bond acceptor features, hydrophobicity, and steric factors may play crucial roles for anti-aromatase activity. Electrostatic, van der Waals, and π-π interactions are other important factors that determine binding affinity of inhibitors. HQSAR, LDA-QSAR, GQSAR, CoMFA, and CoMSIA approaches, pharmacophore mapping followed by virtual screening, docking, and dynamic simulation may be effective approaches for designing new potent anti-aromatase molecules.

INTRODUCTION

Breast cancer, one of the commonest form (accounting for 35% of all cancers) among different types of life threatening malignancies in females, is responsible for 20% of all cancer deaths (Bandi, 2010). More than 5,22,000 women across the world died as a result of breast cancer (May, 2014). The maximum incidence of breast cancer is observed in the Western Europe, North America, Australia and New

DOI: 10.4018/978-1-5225-0549-5.ch004

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Zealand. The incidence of breast cancer is seven fold higher in developing nations. Belgium has the age-standardized highest rate of incidence (more than 110 cases per 1,00,000 women per annum). Apart from that, among top 12 countries, nine belong to Western European, but the Bahamas, Barbados and the United States of America are also in the top ranking. The 12 lowest incidence countries belong to mainly sub-Saharan Africa, South Asia and the far East, those suffer from poverty (May, 2014). The distribution pattern of the age-standardized mortality rate is different across the world. Belgium has the highest mortality followed by the Republic of Ireland with the highest rate of diagnosis but they are outranked by Fiji, Bahamas, Nigeria and Pakistan. Though the mortality is relatively low in low-incidence countries but the mortality rate and probability are higher than the high-incidence countries due to social and cultural influence, stage of presentation and the standards of health care. Due to the high incidence rate, breast cancer ranks top among women's health concerns. Despite the advancement of new preventive strategies against breast cancer consideration, the incidence of breast cancer has remained the same since 2005 (Arumugam *et al*., 2014; Siegel *et al.*, 2012). Breast carcinoma is most frequently diagnosed cancer in women apart from cancer of skin. Approximately 70% of the breast cancers are diagnosed in postmenopausal women (Howlader *et al*., 2014). It ranks second in tumor-related deaths after lung cancer (Muftuoglu & Mustata, 2010). It is predicted that one in eight American women is susceptible to develop invasive breast cancer in their lifetime (American Cancer Society Cancer Facts & Figures, 2013; Brueggemeier, Hackett, & Diaz-Cruz, 2005).

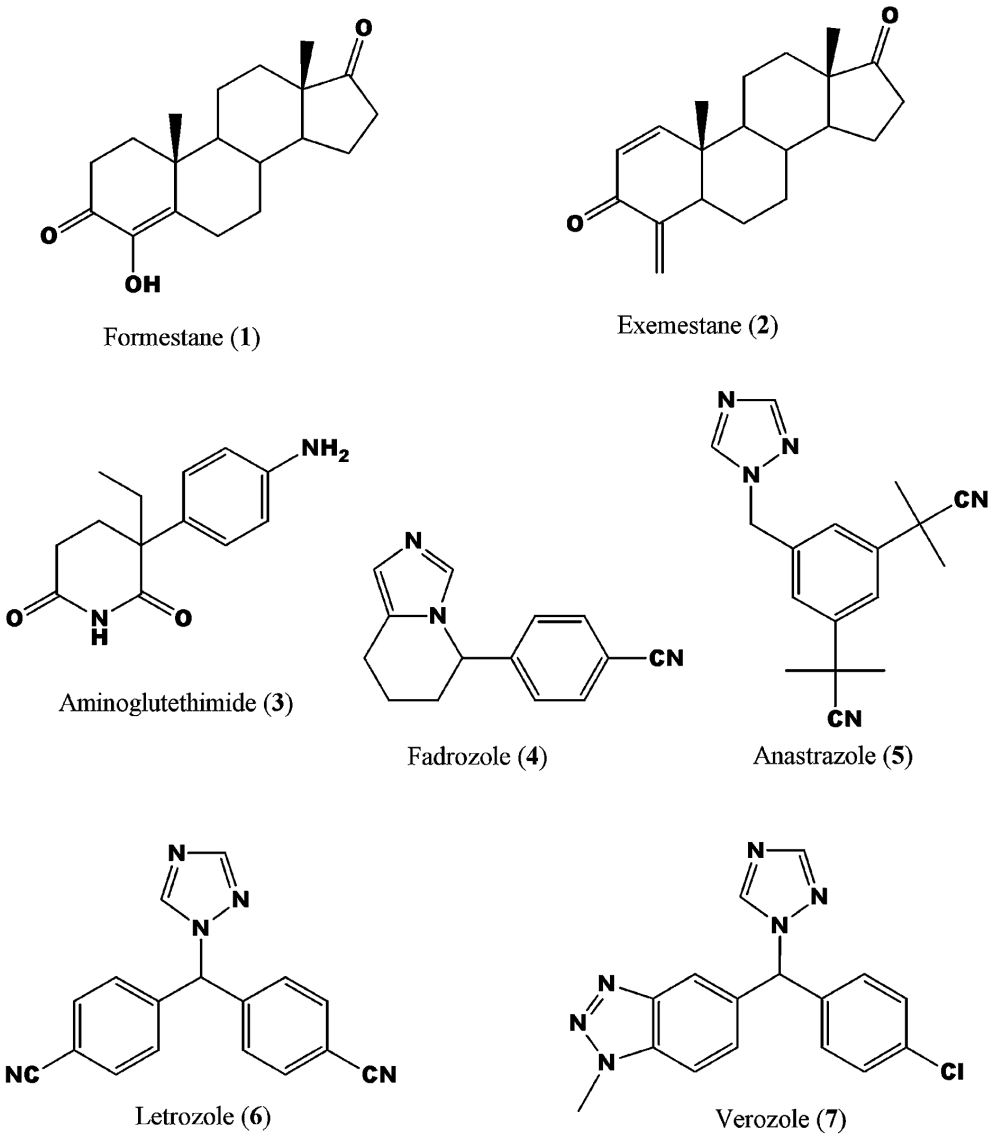
Sex hormones play crucial roles like growth regulation, maturation and reproduction in living animals. Sex hormones are composed of a steroidal cyclopentanoperhydrophenanthrene nucleus. Steroidal hormones are involved in regulation of different physiological effects, like muscle and hair growth, fertility, water retention and dilatation of the capillary vessels as well as sebaceous gland activity (Proteau, 2011; Davis *et al*., 2004; Morales *et al*., 2004). Estrogen is a prominent regulator of cell proliferation in the tumorogenesis of hormone-dependent breast cancer and other tumors. The exact mechanisms of this incidence are still hypothesized. Estrogen metabolites like reactive quinones may directly interact with DNA. It causes mutations that are responsible for these proliferative effects (Miller, 2003). Nearly, two-thirds of breast tumors are hormone-dependent and require estrogens to grow (Brueggemeier, Hackett, & Diaz-Cruz, 2005; Howell, 2005; Muti, Rogan, & Cavalieri, 2006). This phenomenon demands application of an endocrine therapy with a more favorable activity profile and less adverse effects compared to unspecific chemotherapy. More than 100 years ago, Beatson reported that ovariectomy in premenopausal women with breast cancer can induce tumor remission (Beatson, 1896). By knowing the advantage of lowering estrogen level in breast cancer, antihormonal therapy led to the development of new drug candidates. As high serum levels of estrogen is seen in progression of breast cancer, two pharmacological strategies have been employed successfully to control breast cancer (Murthy, Rao, & Sastry, 2004). Drugs either act through estrogen receptor (ER) modulation (Ariazi *et al*., 2006) or interfere with the biosyntheses of steroidal hormones by inhibiting the enzyme controlling the interconversion from androgenic precursors, i.e., aromatase inhibitors (AIs) (Brueggemeier, Hackett, & Diaz-Cruz, 2005). Considerable research work has been devoted to the study of this aromatase enzyme. This helps to develop potent and selective agents that are able to interfere with enzymatic action. Selective estrogen receptor modulators (SERMs) and aromatase inhibitors (AIs) are successfully utilized as therapeutically important weapons in the battle against breast cancer deaths (Pasqualini, 2004). Several classes of steroidal and nonsteroidal aromatase inhibitors are developed (Brodie, Sabnis, & Jelovac, 2006; Neves *et al*., 2009; Colozza *et al*., 2008; Dutta & Pant, 2008; Eisen *et al*., 2008; Gobbi *et al*., 2008; Jackson *et al*., 2008; Osborne & Tripathy, 2005; Osborne & Schiff, 2005; Recanatini & Cavalli, 1998; Recanatini,

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Cavalli, & Valenti, 2002; Spinelli *et al*., 2008). Steroidal aromatase inhibitors (Figure 1) like formaestane (**1**) and exemestane (**2**), and nonsteroidal aromatase inhibitors, e.g., aminoglutethimide (**3**), fadrozole (**4**), anastrozole (**5**), letrozole (**6**) and verozole (**7**) were evaluated successfully in clinical trials (Miller, 2003; Recanatini, Cavalli, & Valenti, 2002).

Steroidal aromatase inhibitors (SAIs) are analogs of androgenic substrates. These inhibit aromatase irreversibly by either competitive or mechanism-based way. Competitive inhibitors bind covalently to the aromatase enzyme, whereas mechanism-based inhibitors, also called suicide inhibitors are converted into a reactive intermediate which covalently binds with the enzyme and thereby permanently inactivates it. The enzyme is often destabilized by this process and the rate of degradation by the intracellular proteosome is increased (Miller *et al*., 2008; Hong *et al*., 2007). On the other hand, nonsteroidal aromatase inhibitors

Figure 1. Aromatase inhibitors (AIs) evaluated in clinical trials



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

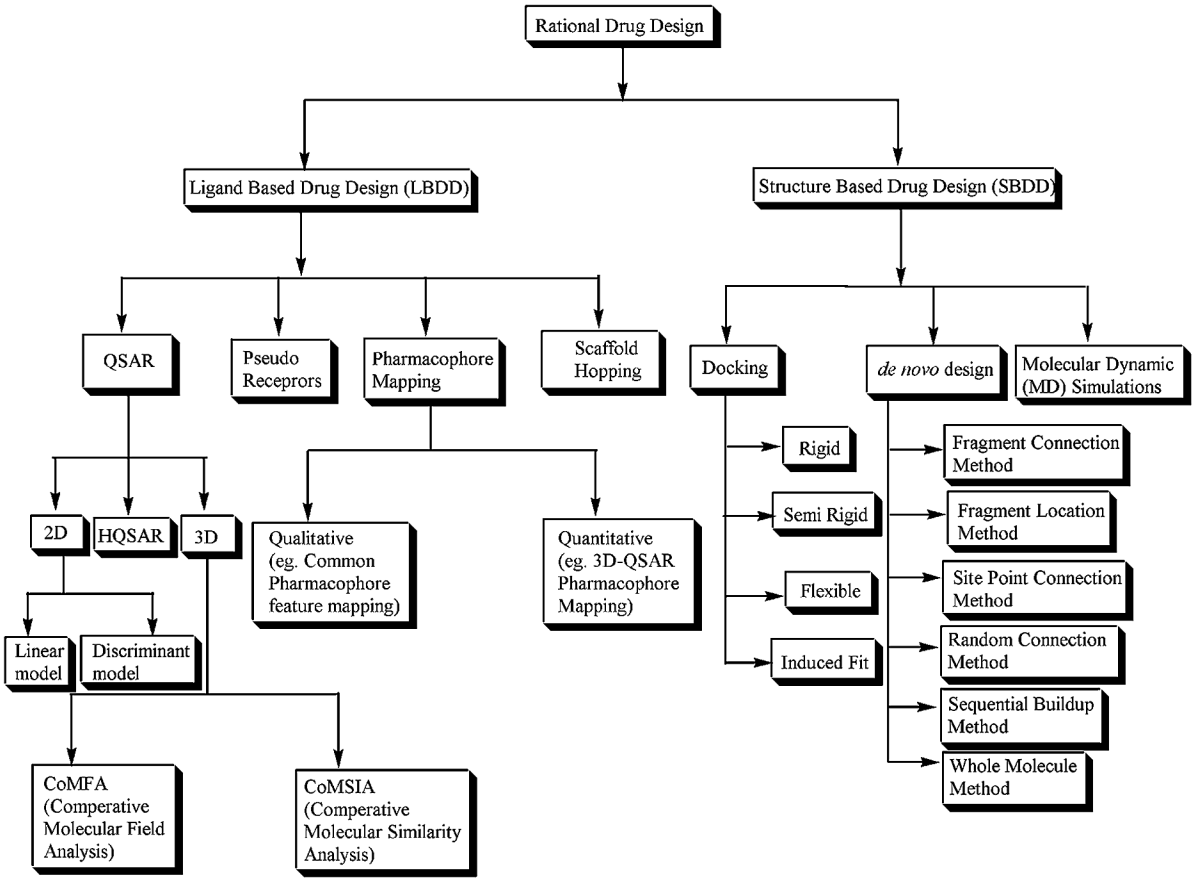
(NSAIs) bind noncovalently to the aromatase enzyme. These block the active site and reversibly inhibit the enzyme (Brueggemeier, Hackett, & Diaz-Cruz, 2005; Hong *et al*., 2007). The NSAIs are divided according to the order of their discovery. Currently, 3rd generation of triazole derivatives is approved as front-line therapy for early and advanced cases of breast cancer in postmenopausal women (Hong *et al*., 2007; Neves *et al.*, 2009). However, as far as the adverse effects are concerned in both steroidal and nonsteroidal AIs, these vary from mild to severe and short term to long term. The SAIs produce androgenic side effects where different other hormone-dependent physiological systems are disturbed due to lack of inhibitor specificity (Geisler & Lonning, 2005). Prolonged estrogen deprivation may lead to osteoporosis and infertility as well as other types of cancers (Hong *et al*., 2007). Hence, benefits of these SAIs may be suppressed due to life-threatening adverse effects. Many patients are compelled to discontinue the use of this type of inhibitors. Thus, more selective and less toxic aromatase inhibitors are needed. These are required as the stability, efficiency and sensitivity to different classes of AIs may vary from patient to patient due to mutations of intratumoral aromatase (Miller *et al*., 2008). The major advantage of NSAIs is devoid of different steroidal adverse effects shown by steroidal AIs.

Drug design, development and discovery are expensive as well as time consuming processes. Traditionally, drug discovery relies on synthesis and screening of large number of compounds to identify a potential lead. Over the decades, there is an increased effort to apply computational approaches to combine both chemical and biological spaces in order to streamline drug design, optimization, discovery and development (Kapetanovic, 2008). Computational methods may play crucial roles in understanding the specific molecular recognition events of the target macromolecule with candidate hits (Shaikh *et al*., 2007). Rational drug design (RDD) methods may reduce time and cost involved in drug development process in comparison to traditional drug discovery methods. The RDD methods are utilized to design new inhibitors as well as for optimizing the pharmacokinetic and toxicity profiles of lead molecules. The RDD approaches (Figure 2) can be categorized in two types, i.e., ligand based drug design (LBDD) and structure based drug design (SBDD) (Aparoy, Reddy, & Reddanna, 2012).

Though a limited number of molecular modeling studies for designing the SAIs (Numazawa, Shelangouski, & Nagasaka, 2000; Beger *et al*., 2001; Numazawa *et al.*, 2002; Polanski & Gieleciak, 2003; Murthy *et al.*, 2006; Dai *et al*., 2010; Roy & Roy, 2010a) and NSAIs (Muftuoglu & Mustata, 2010; Recanatini & Cavalli, 1998; Furet *et al*., 1993; Nagy, Tokarski, & Hopfinger, 1994; Cavalli *et al.*, 2000; Chen *et al*., 2004; Leonetti *et al*., 2004; Cavalli *et al*., 2005; Schuster *et al*., 2006; Nagar *et al*., 2008; Castellano *et al*., 2008; Nagar *et al*., 2009; Karkola & Wahala, 2009; Petkov *et al*., 2009; Nagar & Saha, 2010a; Nagar & Saha, 2010b; Nagar & Saha, 2010c; Roy & Roy, 2010b; Sun *et al*., 2010; Froufe, Abreu, & Ferreira, 2011; Narayana *et al*., 2012; Nantasenamat *et al*., 2013) were performed, structure based as well as ligand based molecular modeling techniques can be used to identify or discover new potential leads with less adverse effects in future that may be beneficial as far as the breast cancer mortality rate is concerned. The chapter highlights the LBDD as well as SBDD studies of NSAIs that can be utilized to design novel anti-aromatase leads useful for the treatment of breast carcinomas. Before discussing the various modeling strategies, the overview of breast cancer and its treatment, and chemistry and functionalities of aromatase enzyme for development of anti-aromatase drugs against breast cancer have been highlighted.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 2. Rational drug design approaches



BREAST CANCER AND ITS TREATMENT

Various factors, such as age, family history, early menarche, late menopause, postmenopausal obesity, use of estrogen and progestin menopausal hormones, alcohol and physical inactivity may be responsible for developing breast cancers. Inheritance of an inactivating mutation in one of the familial breast cancer genes, like BRCA1, BRCA2, CHEK2, p53 and ATM may be the strongest risk factor for the breast cancer, contributing 5% of the breast cancer cases. The second highest risk factor is age. Three-quarter of breast cancer cases (< 5%) is observed in postmenopausal woman less than 40 years of age. But the highest incidence of mortality is observed in women between the age of 34 and 54. The other risk factors include exposure of estrogens, absence of lactation, hormonal therapy and use of oral contraceptive pills (May, 2014; Westley and May, 2013; McCullough *et al.*, 2012; Allen *et al.*, 2009). Nevertheless, the risk of breast cancer may be observed in diabetic conditions like high circulating insulin-like growth factor-1 (IGF-1) and low serum insulin-like growth factor binding protein 3, though exercise may reduce the risk of breast cancer. The chance of developing invasive breast cancer is strongly influenced by the extent of spread of cancer when it is first diagnosed. The American Joint Committee on Cancer (AJCC) classified tumors on the basis of tumor size and the propagation within the breast and nearby organs (T) and involvement of lymph node (N) as well as the presence or absence of different metastases in distant

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

organs (M) (AJCC Cancer Staging Manual, 2002). If T, N and M are determined, stages of I, II, III, and IV are assigned, where stage I is an early stage and stage IV is the most advanced.

The early detection of breast cancer uses mammography, clinical breast examination (CBE) and magnetic resonance imaging (MRI). Mammography may be able to detect about 80-90% of the breast cancer (Michaelson, Satija, & Moore, 2002). For women aged 40 and above, annual CBE may be an important complement to mammography. A small percentage of breast cancers may be missed during mammography can be detected in CBE. The shape, texture, location of lumps and if any skin color changes are observed during CBE. In MRI, magnetic fields are used instead of X-ray mammography. It generates detailed cross-sectional images of breast tissue. As far as the breast cancer treatment is concerned, it includes both local and systemic therapy to get rid of tumor as completely as possible and to prevent its recurrence. Local therapy includes surgery and radiation therapy. In lumpectomy, only cancerous tissues along with some portion of normal tissues are removed. In mastectomy, removal of complete breast occurs. Lumpectomy followed by radiation may affect the same expected long time survival rate as in mastectomy (Fisher, Anderson, & Bryant, 2002). In both lumpectomy and mastectomy, removal of axillary lymph nodes is done to determine the spreading beyond breast. The presence of cancerous lesions in lymph node helps to determine subsequent therapy. Sentinel lymph node biopsy reduces the need for full axillary lymph node dissections in majority of women with no evidence of lymph node enlargement before therapy (Lyman, Giuliano, & Somerfield, 2005). Radiation therapy destroys cancerous cells after or before surgery in the breast, chest wall and or underarm area (Early Breast Cancer Trialists' Collaborative group, 2000). Radiation therapy depends on the type and stage as well as location of the tumor being treated. Apart from the local therapy, systemic therapy includes biological therapy and chemotherapy as well as hormone therapy. Systemic treatment with anticancer drugs prior to surgery is a neoadjuvant therapy. This helps in shrinkage of tumor which relieves the stress of surgical removal. In adjuvant therapy, a systemic treatment is given to the patients after surgery. This therapy is more effective in terms of survival, disease progression and distant recurrence (Mauri, Pavlidis, & Loannidis, 2005).

Approximately, about 15-30% of breast cancers are related to overproduction of the growth promoting protein 'Human Epidermal Growth Factor Receptor 2' (HER2/neu). It is also known as ErbB2 protein. It gives higher aggressiveness in breast cancers. It belongs to ErbB2 protein family or epidermal growth factor receptor (EGFR) family. It is also known as CD340 (cluster of differentiation 340) and p185. It is encoded by ErbB2 gene (Gligorov & Lotz, 2008). Overexpression of this receptor is associated generally with increased disease recurrence as well as worse prognosis. Since the prognostic role of the receptor and its ability to predict response against trastuzumab (herceptin), breast tumors are checked regularly for overexpression of HER2/neu [since the oncogene 'neu' is derived from a rodent gliboblastoma cell line (a neural tumor), it is termed as 'neu']. HER2 is named on the basis of its similar structure to human epidermal growth factor receptor 1 (HER1). Similarly, ErbB2 is termed for its similarity to ErbB (avian erythroblastosis oncogene B). It is found to encode EGFR. Gene cloning study showed that neu and HER2 as well as ErbB2 are same protein. The HER2 is co-localized and therefore, co-amplified with the gene GRB7. The GRB7 is a proto-oncogene found active in breast, gastric and esophageal cancers. Interestingly, it is found that patients with $ER^+/HER2^+$ get more benefit from drugs compared with $ER/HER2^+$ breast cancers by inhibiting PIK3/AKT pathway. The HER2/neu is an important target for monoclonal antibody trastuzumab (herceptin) (Piccart-Gebhart *et al*., 2005). Herceptin is effective only in breast cancer when HER2/neu receptor is overexpressed. Overexpression of HER2 gene may be suppressed by the amplification of other genes as well as use of trastuzumab. Another monoclonal antibody, pertuzumab is in clinical trials. It can inhibit the dimerization of HER2 and HER3 receptors. The

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

expression of HER2/ErbB2 protein is generally regulated by ERs. Again, estradiol and tamoxifen acting through ERs are found to downregulate HER2/ErbB2 expression. When the coactivator AIB-3 exceeds the corepressor PAX2, the HER2/ErbB2 is upregulated in presence of tamoxifen. It leads to tamoxifen resistant breast cancer (Hurtado *et al*., 2008; Vogel, Cobleigh, & Tripathy, 2002; Salmon, Leyland-Jones, & Shak, 2001). Lapatinib is responsible for delay in disease progression in $HER2^+$ breast cancer resistant to trastuzumab. New generation anti-$HER2^+$ targeted therapies are in development (Bedard, Azambuja, & Cardoso, 2009). Bevaxizumab retards metastasized tumor growth by blocking angiogenesis (Early Breast Cancer Trialists' Collaborative Group, 2005).

As far as the molecular targeted therapy of breast cancer is concerned, there are several molecular targets, and inhibiting the functionality of the targets is effective for treating breast cancer, like HER-2 inhibitors (Brufsky, 2014), kinase inhibitors of phosphatidyl-inositol 3-kinase (PI3K), protein kinase B (Akt), mammalian target of rapamycin (mTOR) and tyrosine, IGF1R inhibitors, cyclin dependent kinase inhibitors (CDKs), heat shock protein 90 (HSP90) inhibitors, telomerase inhibitors, vascular endothelial growth factor (VEGF) inhibitors, poly-(ADP-ribose) polymerase (PARP) inhibitors, inhibitors of apoptosis proteins (IAPs) (Schlotter et al., 2008), nuclear transcription factor (NF-κB), ubiquitine-proteasome system and farnesyl transferase as well as p53 targeted gene therapy may be effective for development of anti-breast cancer drugs (Schlotter et al., 2008). Apart from that, selective estrogen receptor modulators (SERMs) and aromatase inhibitors (AIs) play vital role in prevention of breast cancer. The IAPs prevent uncontrolled and excessive cell death in apoptotic signaling. Survivin, supposed to be a possible target of breast cancer, is detected in almost 90% of breast tumors (Schlotter *et al*., 2008). Increased level of survivin is found to be related in HER2 overexpression, VEGF overexpression and high urokinase type plasminogen activator levels (Schlotter *et al*., 2008; Ryan *et al*., 2006) It plays a key role in multi-drug resistance (MDR) in presence of p-glycoprotein, and this may be a novel strategy for modulating MDR resistant cancer cells (Liu et *al.*, 2007). Activation of NF-κB may be responsible for progression of hormone dependent breast cancer (Nakshatri *et al.,* 1997). The NF-κB is also found to inhibit extracellular signal-regulated kinase (ERK) activation to enhance cell survival during the development of tumour adaptive radioresistance (Ahmed *et al*., 2006). The ubiquitine-proteasome system was reported to regulate the p53 cyclins, cyclin-dependent kinases (CDKs) and the bcl family proteins. Inhibitors of the proteasome system may be responsible for the accumulation of pro-apoptotic proteins, which deliver cytochrome from mitochondria and thereby activating intrinsic apoptotic signal transduction (Schlotter *et al.*, 2008). Inactivation of NF-κB by proteasome inhibition may be responsible for increased apoptosis by histone deacetylase (HDAC) inhibitors induction. The HDAC inhibitors are a novel class of anti-cancer agents, play important roles like cell growth arrest, cellular differentiation and apoptosis in many tumour cells. They may also regulate the activity of the anti-apoptotic transcription factor NF-κB (Domingo-Domenech et al., 2007). The PI3K is reported to play an important role in survival, proliferation, motility and neoangiogenesis in cancer cells. Due to the disregulation of the cell cycle, PI3K is overexpressed and may activate Akt. By means of phosphorylation of NF-κB, bad and caspase 9, Akt may have anti-apoptotic influences. The EGFR and HER2 may also activate PI3K and responsible to enhance cell growth via Akt. Therefore, PI3K may be a valuable target for the breast cancer (Schlotter et al., 2008). The mTOR protein is activated by phosphorylated Akt. Hence, mTOR protein is found to promote cancer cell proliferation and neoangiogenesis. It may also promote tumor spread and metastasis and inhibit cancer cell apoptosis. Inhibition of mTOR by temsirolimus may have negatively influences on these mechanisms, investigated in clinical trials (Schlotter et al., 2008). The Hsp90, a chaperone protein, is reported to promote protein folding and stabilization, hence prevent rapid protein degradation. The

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Hsp90 inhibitors, like tanespimycin, retaspimycin and ganetespib are proved to be beneficial against HER2 resistant breast cancer in clinical trials. However, the clinical study for tanespimycin was terminated because of corporate decision, whereas the investigation was discontinued due to lack of efficacy (Modi *et al*., 2011; Modi *et al.*, 2013). The expression of telomerase is observed in cellular proliferation that has a relation with tumorogenesis. Inhibition of telomerase may cause apoptosis (Goldblatt *et al*., 2009). Telomerase inhibitor, imetelstat is reported to restore trastuzumab sensitivity in trastuzumab resistant cell line. However, phase 1 study of imetelstat showed no response in metastatic breast cancer (Miller, 2012). A crosstalk between IGF1R and HER2 in breast cancer may lead to heterodimerization and may cause tumor cells to escape trastuzumab toxicity. However, in clinical trials, it was proved that IGF1R inhibitor and trastuzumab combination resulted better activity than trastuzumab alone ((Nahta *et al*., 2005). Treatment with linsitinib, an IGF1R inhibitor, in metastatic breast cancer was discontinued in phase 2 study due to lack of efficacy and severe toxicity (Brufsky, 2014). Cyclooxygenase-2 (COX-2) inhibitors were also a promising target for breast cancer. Celecoxib was found to be effective in ER-negative breast cancer and triple negative breast cancer (TNBC) by delaying tumor progression (Howe *et al*., 2002). Unfortunately, due to increased risk of heart failure,clinical trials of COX-2 inhibitors were halted by FDA though there were positive breast cancer prevention clinical trials (Hollander *et al.*, 2013).

Chemotherapy helps to stop the growth of cancer cells either by killing or by arresting their multiplication. It is dependent on the breast tumor size, number of lymph nodes involved, presence of estrogen and/or progesterone receptors as well as the amount of HER2/neu protein of cancer cells. It is observed that combination of drugs (combination chemotherapy) is more effective than single drug for breast cancer treatment (Hortobagyi, 2006; Mauri *et al*., 2008). The common drugs used in combination are florouracil, methotrexate, doxorubicin, cyclophosphamide, paclitaxel and docetaxel.

Hormonal therapy is useful as it blocks the effect of estrogen in breast cancer cells. Anti-estrogenic drugs act either by blocking the hormone receptor or destroy ovaries or suppress the estrogen production. The growth stimulatory effects of estrogens may be reduced either by using SERMs or by using AIs. The SERMs may compete for binding to ERs (competitive inhibition) and reduce the number of ERs available for binding to endogenous estrogen. This approach is an effective anticancer strategy. It leads to the development of antiestrogenic drugs, such as tamoxifen, raloxifene and tormifene (Carmichael, 1998; Jordan, 1995). Inhibition of aromatase enzyme is most effective approach to reduce growth stimulatory effects of estrogen.

AROMATASE BIOCHEMISTRY

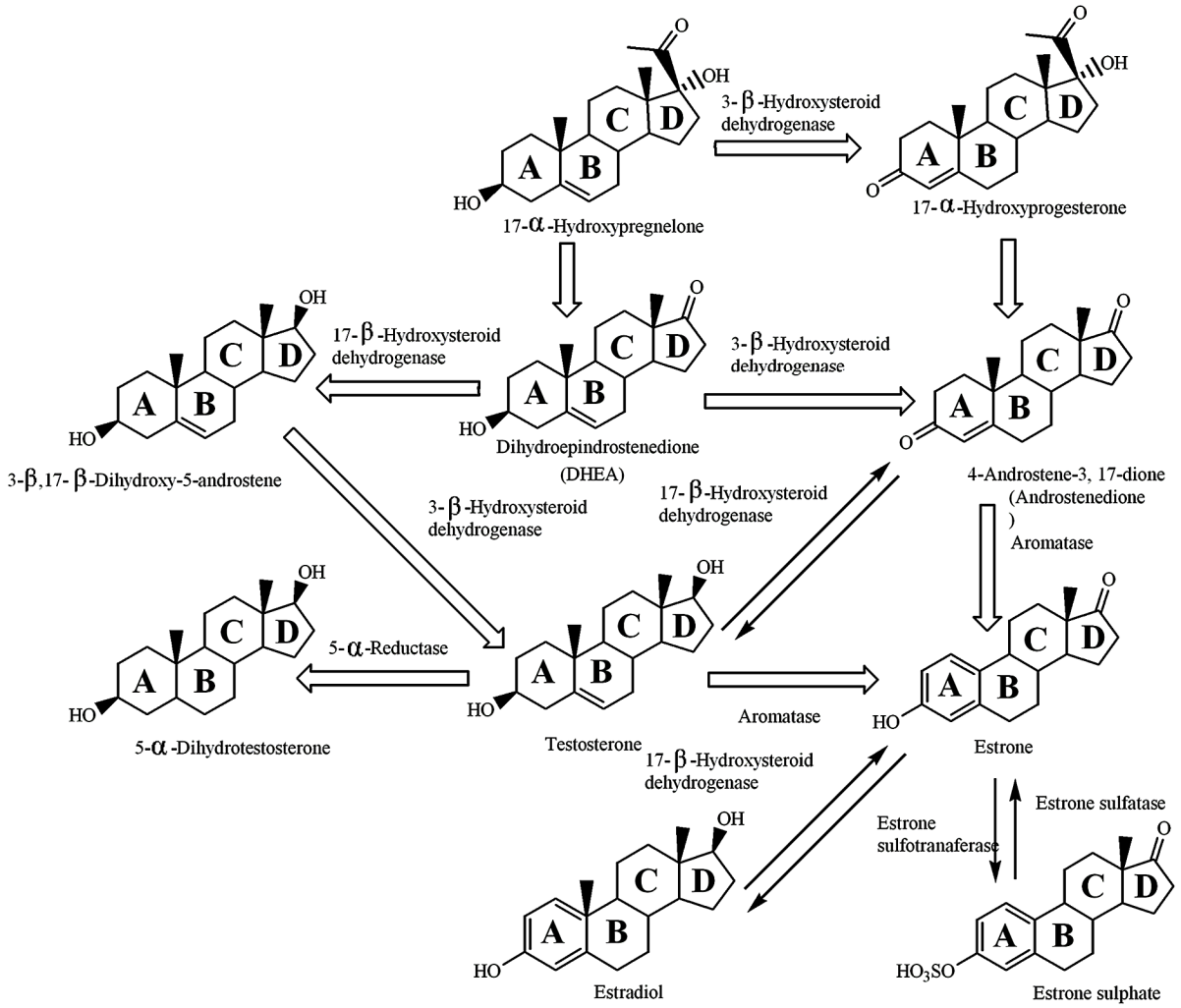
Effective aromatase inhibitors (AIs) are developed as the therapeutic agents for estrogen-dependent breast cancer. Development of novel AIs started in early 80s of the last century and continued greatly in last four decades (Brueggemeier, Hackett, & Diaz-Cruz, 2005; Murthy, Rao, & Sastry, 2004; Recanatini, Cavalli, & Valenti, 2002; Geisler & Lonning, 2005; Harvey, Lipton, & Santen, 1982; Covey, 1988; Johnston & Metcalf, 1984; Banting *et al*., 1988; Banting *et al*., 1989; Cole & Robinson, 1990; Brueggemeier, 1990; Brueggemeier, 1994; Brodie & Njar, 1996, 2000; Brodie, Lu, & Long, 1999; Santen & Harvey, 1999; Seralini & Moslemi, 2001; Geisler, 2003; Smith & Dowsett, 2003; Altundag & Ibrahim, 2006; Benson & Ravisekar, 2007). Estrogen biosynthesis in postmenopausal women occurs principally in extragonadal sites where adipose tissue plays an important role. The peripheral estrogen production is important. This may be the reason for occurrence of breast tumors in postmenopausal women. Furthermore, it is observed

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

that 17β-estradiol may be produced in breast tissue, and thereby estrogen concentration may be maintained at an appropriate level for stimulating breast cancer growth. Even biosynthesis of small amount of estrogen may cause breast tumor development (Labrie *et al*., 1997). After menopause, biologically important estrogen, i.e., 17β-estradiol is generated from the circulating androstendione from adrenal gland. Androstenedione is transformed to estrone. This is converted into estrone sulfate as a reservoir of estrogen hormone. Estrone may be derived from the aromatization of androgen or from the sulfate metabolite. After removal of the sulfate group, it is converted to 17β-estradiol. Aromatase catalyzes the syntheses of estrogens. This is done via the aromatization of the A ring of androgen precursors, namely androstenedione and testosterone. Aromatization is considered to be responsible for the biosyntheses of estrogens from androgens (Figure 3) (Schuster *et al*., 2006).

Estradiol is the most important endogenous estrogen which is biosynthesized from androgens by the cytochrome P450 enzyme complex known as 'aromatase' (Simpson *et al*., 1994). Aromatase is a multienzymatic complex which is mostly expressed in the ovary of premenopausal women, placenta,

Figure 3. Biosynthetic pathway of estrogens and androgens



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

peripheral adipose tissues, breast tissue and in the brain (Simpson *et al*., 1994). It is overexpressed in breast cancer tissue and may be responsible for local estrogen production and proliferation of breast tumors (Millet *et al*., 2008; Miller, 1997; James *et al*., 1987; Miller & O'neill, 1987; Reed *et al*., 1989; Bulun *et al*., 1993). Adipose stromal cells surrounding the cancer cells are also reported to contain higher aromatase mRNA level than adipose stromal cells in noncancerous areas (Dickson & Lippman, 1995; Harada, 1997). Again, estrogen concentration in breast cancer tissue is found to be several-fold higher than plasma in postmenopausal women. This is supported by the synthesis of estrogens with the aromatase enzyme (Pasqualini *et al*., 1996). Out of 70 breast tumor specimen, aromatase mRNA is also found in 67 specimens (Zhou *et al*., 1996). The presence of aromatase in breast cancer epithelial and stromal cells is also found by immunocytochemical analysis (Esteban *et al*., 1992). Gene transcriptional studies reveal that aromatase promoter switches from a glucocorticoid stimulated promoter I.4 in normal tissue to cAMP stimulated promoters I.3 and II in cancereous tissue. Prevention of aromatase expression suppresses *in situ* estrogen biosynthesis. Anti-aromatase drugs are a second-line treatment in patients who fail anti-estrogen therapy. The patients (20-30%) failed to anti-estrogen therapy, may respond to aromatase inhibitors (Chen, 1998). Aromatase enzyme complex is located in the endoplasmic reticulum of cells. This is composed of a cytochrome P450 heme protein (CYP19). The CYP19 carries out the aromatization reaction. Simultaneously, a NADPH-cytochrome P450 reductase (a flavoprotein) is required for the electron transfer from NADPH to the cytochrome P450 enzyme (Brodie & Njar, 1998; Simpson & Davis, 2001). The heme protein (CYP19) is responsible for converting the alicyclic A ring of C19 steroids (androgens) to C18 steroids (estrogens) containing the phenolic A ring (Kellis & Vickery, 1987), whereas NADPH-cytochrome P450 reductase is responsible for transferring reducing equivalents to cytochrome P450 (CYP19). In the three step process, aromatase converts androstendione, testosterone and 17α-hydroxytestosterone to estrone, 17β-estradiol and 17β, 16α-estriol, respectively by coupling with 1 mol O_2, 1 mol NADPH and cytochrome P450 reductase (Simpson *et al*., 1994). The first two steps involve C19 methyl hydroxylation. The third step is responsible for the aromatization of the ring A which is a unique characteristic of aromatase enzyme.

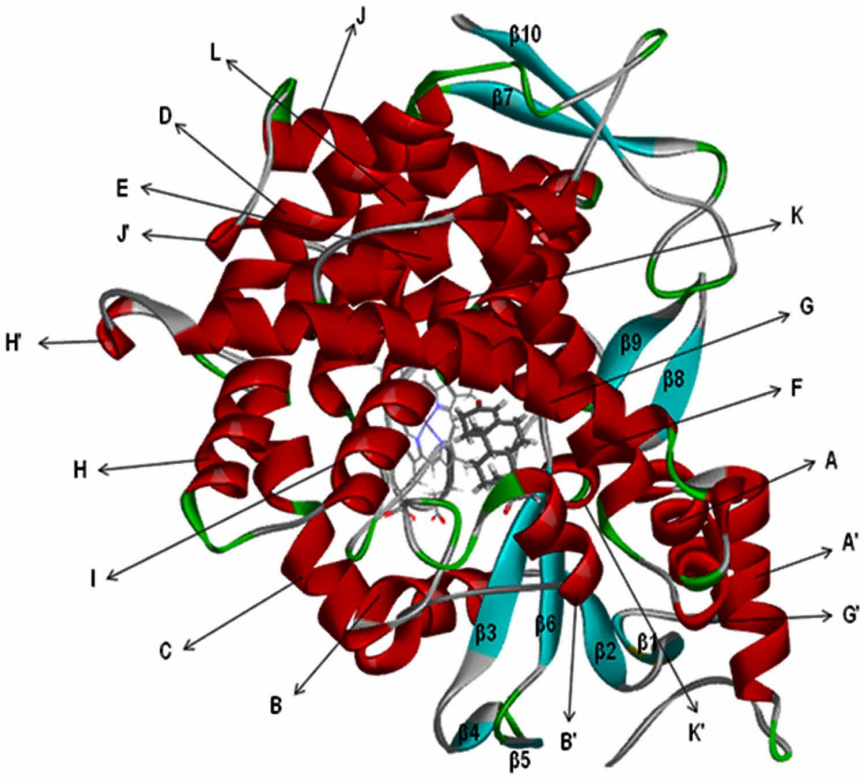
STRUCTURE OF AROMATASE

Aromatase is the only known enzyme in vertebrates capable of catalyzing the aromatization reaction of a six-membered alicyclic ring. The functional human aromatase is monomeric. It is comprised of a heme group as well as a single polypeptide chain of 503 amino-acid residues. It is an integral membrane-bound protein of the endoplasmic reticulum. It is attached to the membrane by an amino-terminal transmembrane domain, in addition to other membrane-associating regions (Shimozawa *et al*., 1993; Amarneh *et al.*, 1993). It shows high substrate specificity for syntheses of estrogens from androgen precursors. The crystal structure of placental aromatase showed key aspects regarding the properties of aromatase (Ghosh *et al*., 2009, 2010, 2011; Jiang & Ghosh, 2012). A ribbon diagram of the overall crystal structure of human placental aromatase is shown in Figure 4.

The tertiary structure of human aromatase is made up of 12 major α-helices as well as 10 β-strands. These β-strands are distributed into 1 major and 3 minor sheets. These show the characteristic cytochrome P450 fold (Ghosh *et al*., 2010). The amino terminus starts at residue 45, whereas the carboxyl terminus is ending at residue 496. Androstenedione binds at the heme distal site. This is the active site of the enzyme. The 3-keto and 17-keto oxygens of androstenedione are responsible for two H-bond contacts

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 4. The structure of human placental aromatase (ribbon diagram) along with the heme group and the bound androstenedione molecule at the active site; the α-helices are labeled from A to L and β-strands are numbered from 1 to 10.
(Ghosh et al., 2009).



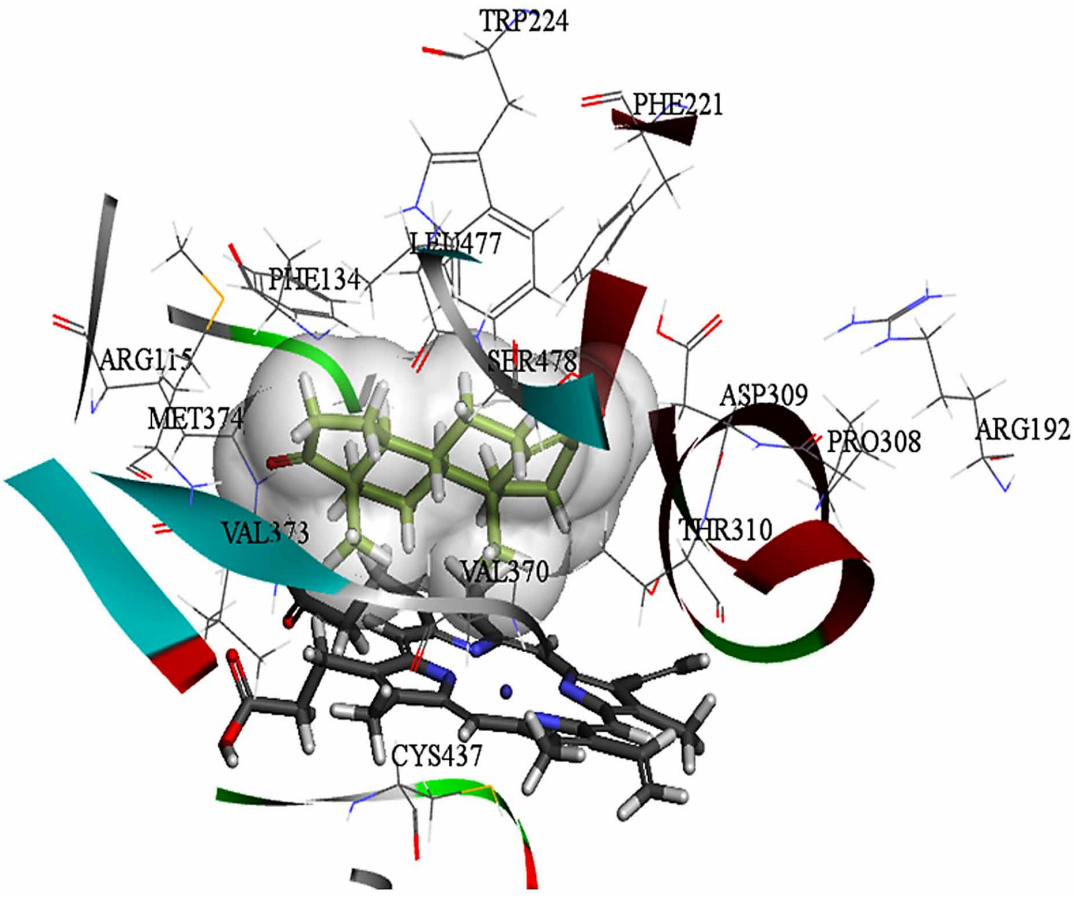
with Asp309 and Met374 respectively. Among the 12 major α-helices, helices I (amino acid residue 293-324), F (residue 210-227), G (residue 242-267), H (residue 278-287), C (residue 138-152), D (residue 155-174), E (residue 187-205), J (residue 326-341), K (residue 354-366) and L (440-455) are similar in most of the cytochrome P450s. Other α-helices, i.e., A′ (residue 57-68), A (residue 69–80), B (residue 100-109), B′ (residue 119-126), G′ (residue 232-236), H′ (residue 271-274), J′ (residue 346-349), K′ (residue 398-404) and K′′ (residue 414-418) are 1-4 turns long. These consist of more variability among P450s in their locations, lengths and orientations (Ghosh *et al*., 2010). The major β-sheet is a 4-stranded sheet. The first two strands begin near the amino terminus (β1:83-88 and β2:93-97). These end in two strands from the carboxyl terminal half of the polypeptide chain (β3:373–376 and β6:393-396) (Ghosh *et al*., 2010). Amino terminal residues 47-50 of the aromatase structure are having one backbone H-bond with β1. Each of the three minor sheets is made up of two anti-parallel strands. These are scattered over the polypeptide chain (sheet2: β4:381-383 and β5:386-388; sheet3: β8:473-475 and β9:479-481; sheet4: β7:458-461 and β10:491-494) (Ghosh *et al*., 2010). Long loops interconnect with the secondary structure in tertiary aromatase structure. Long loops between helices like B′ and C, β7 and β8, and β9 and β10 are either in active site residues or are responsible for participation in the scaffolding of functionally important elements. These loops are having intra-loop H-bonding interaction. This is essential for their structural stabilities. All cytochrome P450s are found to bind the heme group through

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

its propionate moieties. These are done by arginine and tryptophan side chains via ionic and hydrogen bonding. Residues in aromatase are Arg115, Trp141, Arg145, Arg375 and Arg435, homolog to those in 3A4, 2D6 and other proteins (Ghosh *et al*., 2010). The active site of aromatase is having the distal cavity of the heme binding pocket. The heme iron is observed at the reaction center of the enzyme. The active site is buried deep inside the roughly spherical molecule near its geometrical center. Androstenedione binds with its β-face oriented towards the heme group. The C19 methyl group is located at a distance 4.0 Å from the Fe-atom (Figure 5).

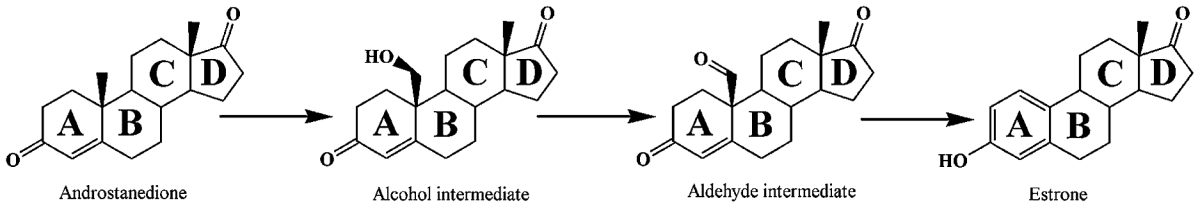
The refined Fe-position is displaced around 0.2 Å away from the heme plane towards the amino acid residue Cys437 of aromatase enzyme. The oxygen atom of 17-keto group of the substrate is located at a distance 2.8 Å from the backbone amide nitrogen of Met374. This helps to accept a proton as well as makes a hydrogen bond. The 3-keto oxygen on the other end is at 2.6 Å away from the carboxylate oxygen of the Asp309 side chain. This suggests protonation of the carboxylate moiety and the formation of a hydrogen bond. The geometries of these two hydrogen bonds are in a fashion such that 3-keto oxygen

Figure 5. Close up view of the human placental aromatase active site bound androstenedione molecule within its unbiased electron density surface along with important side chains, heme and water molecules are shown (distances in Å)
(Ghosh et al., 2010).



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

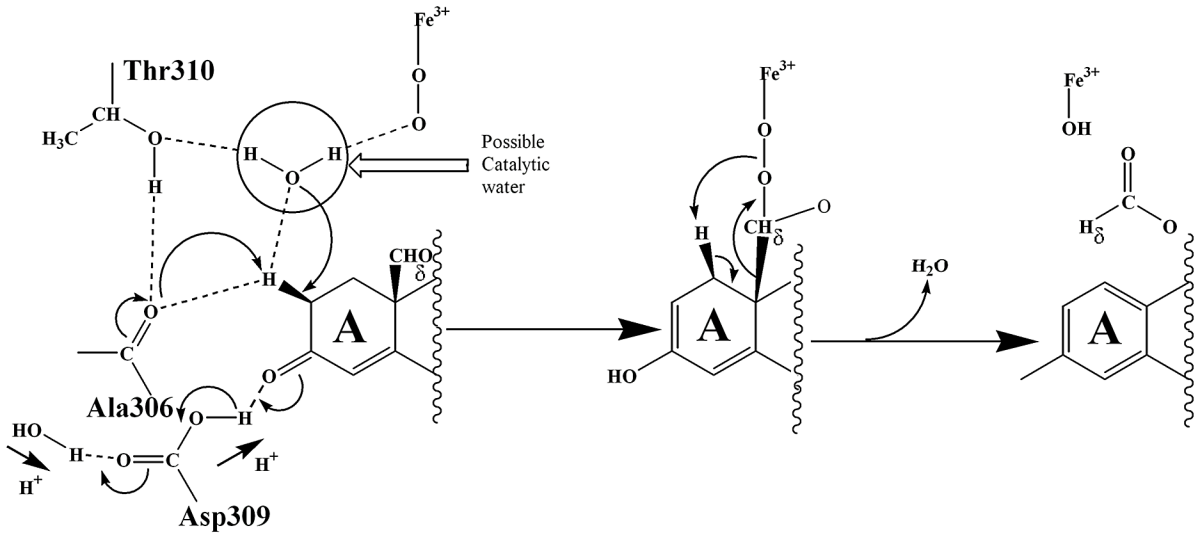
Figure 6. Androstenedione-aromatase interactions (close view of Ala306CO…HOThr310 pair that play key role in aromatization of androstenedione A ring)
(Ghosh et al., 2009).



and the water oxygen atom lie approximately in the carboxylate plane. Again, this water molecule is situated at 3.6 Å from the guanidinium group of the Arg192 side chain. This is salt-bridged to Glu483 (not shown). Two water molecules are found hydrogen-bonded to each other and also to the hydroxyl group of Ser478. This, in turn, donates its proton to His480 nitrogen atom from the active site. Again, the Ser478 side chain is linked through these two water molecules to Arg192 by a weak hydrogen bond of length 3.4 Å (Figure 5). The hydroxylation step involves the aromatization of the A ring via the involvement of Ala306, Thr310, Asp309 and the heme iron of the aromatase enzyme. In the first two steps, the C-19 methyl is transformed into aldehyde derivative of androstenedione (Figure 6) (Ghosh *et al*., 2009).

Thereafter, nucleophilic attack on H2β by the Ala306 CO moiety as well as electrophilic attack on the C3 keto oxygen by a protonated Asp309 side chain help the H2β abstraction by catalytic water molecule (stabilized by hydrogen bond interactions with Thr310, heme iron and H2β of androstenedione) and 2,3-enolization by protonated Asp309 (which is again protonated by bound water molecule) (Ghosh *et al*., 2009) (shown in Figure 7). The aromatization is over when the C-19 aldehyde of androstenedione intermediate is abstracted by ironperoxy residue of heme to form formic acid and water.

Figure 7. Possible mechanism for H2β abstraction and 2,3-enolization of androstenedione
(Ghosh et al., 2009).



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

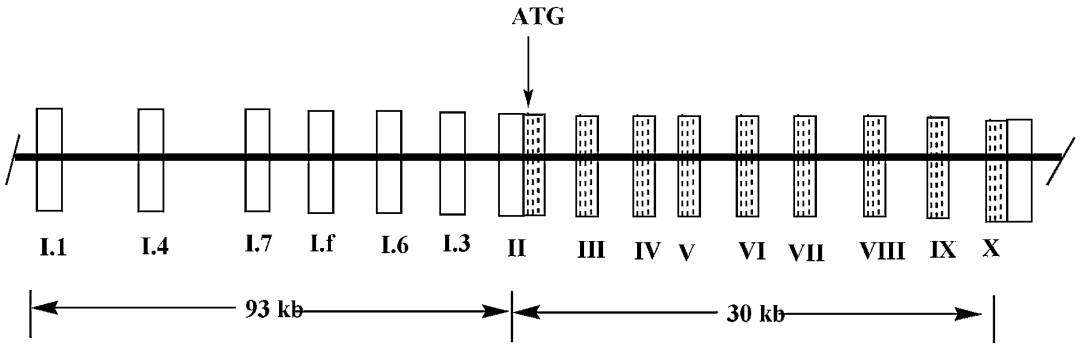
Thus, the crystal structure of human aromatse reveals a finely orchestrated molecular machine that produces estrogen from androgens. The molecular basis of enzyme-substrate and enzyme-drug interactions, more effective intervention of estrogen production may lead to the discovery of potent inhibitors.

EXPRESSION OF GENE OF AROMATASE

The aromatase gene is designated as CYP19. It encodes the cytochrome P450-aromatase. It is found at the chromosome 15q21.1. The coding region is nearly 30 kb in size, whereas the regulatory region is approximately 93 kb (Simpson *et al*., 1993; Bulun *et al*., 2004). Aromatase gene consists of 10 exons. Its full length cDNA (3.4 kb) encodes the protein consists of 503 amino acids. The aromatase enzyme is a glycosylated cytochrome P450 protein having a molecular mass of 58000 Da approximately (Gartner *et al*., 2001). Regulation of aromatase enzyme is a complex mechanism in tissues and tissue-specific promoter regions identified from the CYP19 gene (Simpson *et al*., 1993, 2002; Zhao *et al*., 1997). The tissue specific promoters consist of PI.1, PI.3, PI.4, PI.6 and PI.7 as well as PII. The tissue specific aromatase gene and promoter regions are shown in Figure 8.

Promoter I.1 (PI.1) is one of the major promoters utilized in placental tissues. The PII promoter is utilized in the ovary and breast cancer tissues. It has a cAMP response element. Promoters PI.3, PI.4, PI.6 and PI.7 are utilized in extraglandular sites. The PI.4 is the primary promoter and it is utilized in normal adipose tissue. It is responsive to glucocorticoids and cytokines, e.g., IL-1, IL-6 and TNF. Promoter PI.3 is available in breast adipose tissues. It is elevated in breast cancer.

Figure 8. Tissue specific aromatase gene and promoter regions



Tissue specific promoters:

I.1 Placental tissue
I.2 Normal adipose tissue
I.7 Vascular endothelial tissue
I.f Brain
I.6 Bone
I.3 Adipose tissue; breast cancer
II Ovary; breast cancer

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

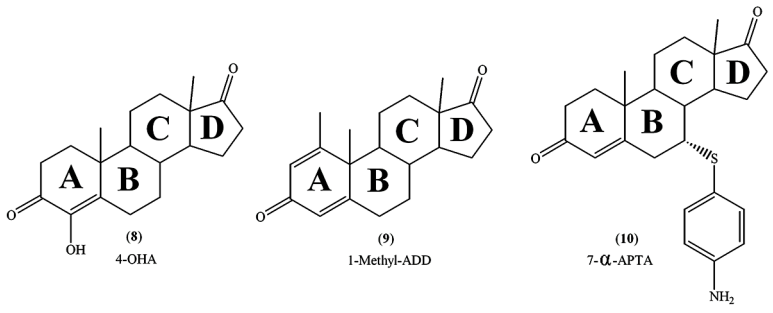
AROMATASE INHIBITORS (AIs)

Aromatase inhibitors are mainly categorized into three generations based on their chronology of clinical trials. These are also classified as type-I and type-II inhibitors depending on their mechanisms of action. Aminoglutethimide (AGT) (**3**) (Figure 1) is the first generation AI, whereas the second generation AIs are formestane (FMS) (**1**) and fadrazole (FDZ) (**4**) (Figure 1). The second generation AIs are more potent than the first generation AIs but these have either poor pharmacokinetic profile or can affect other steroidogenic enzymes. The third generation AIs, such as anastrazole (ANZ) (**5**), letrozole (LTZ) (**6**) and exemestane (EXM) (**2**) (Figure 1) are extremely potent. These are specific in inactivating aromatase at nanomolar concentrations. Again, AIs are also divided chemically into steroidal (SAIs) and non-steroidal (NSAIs) molecules. The use of first generation AIs like AGT (**3**) and second generation AIs like FMS (**1**) and FDZ (**4**) are limited (Ariazi *et al*., 2006), whereas third generation AIs are now used to treat breast cancer. The third generation AI candidates were developed in early 1990s of the last century. These include the triazole derivatives [e.g., ANZ (**5**) and LTZ (**6**)] and steroidal derivatives [e.g., EXM (**2**)] (Smith & Dowsett, 2003). SAIs are again classified into two types depending on the mechanism of action. These are termed as competitive enzyme inhibitors and mechanism-based enzyme inhibitors. Competitive inhibitors compete with androstenedione through non-covalent binding to the aromatase active site and thereby, block its action. Investigations of the AIs started with the synthesis and biochemical evaluation of this type of inhibitors (Schwarzel, Kruggel, & Brodie, 1973; Sitteri & Thompson, 1975; Brueggemeier, Floyd, & Counsell, 1978). Steroidal competitive enzyme inhibitors are developed on the basis of the androstenedione nucleus and the substitutions were done at the varying positions on the steroid (Figure 9).

The mechanism-based inhibitors mimic the substrate and are transformed by the enzyme to reactive intermediates. These result in inactivation of aromatase enzyme. These inhibitors, in other terms are called as 'enzyme-activated irreversible inhibitors' or 'suicide substrates' or 'suicide inactivators'. This type of AIs is also structurally similar to the natural substrate androstenedione (Figure 10).

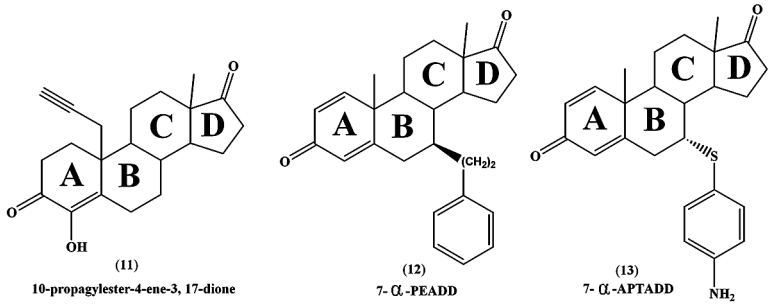
As far as the NSAIs are concerned, these consist of a heteroatom as a common chemical feature. These interfere with steroid hydroxylation by the binding of the heteroatom with the heme iron of aromatase enzyme. AGT (**3**) (Figure 1) is the prototype for NSAIs (Cocconi, 1994) since it is first studied in patients. It should be administered with corticosteroids because of the inhibitory effects on cortisol and aldosterone biosyntheses (Cocconi, 1994; Hoffken, 1993). FDZ (**4**) (Figure 1), an imidazole derivative, acts as a potent second generation inhibitor of aromatase enzyme both *in vivo* and *in vitro* (Trunet *et al*.,

Figure 9. Steroidal competitive enzyme inhibitors



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 10. Steroidal mechanism-based aromatase inhibitors



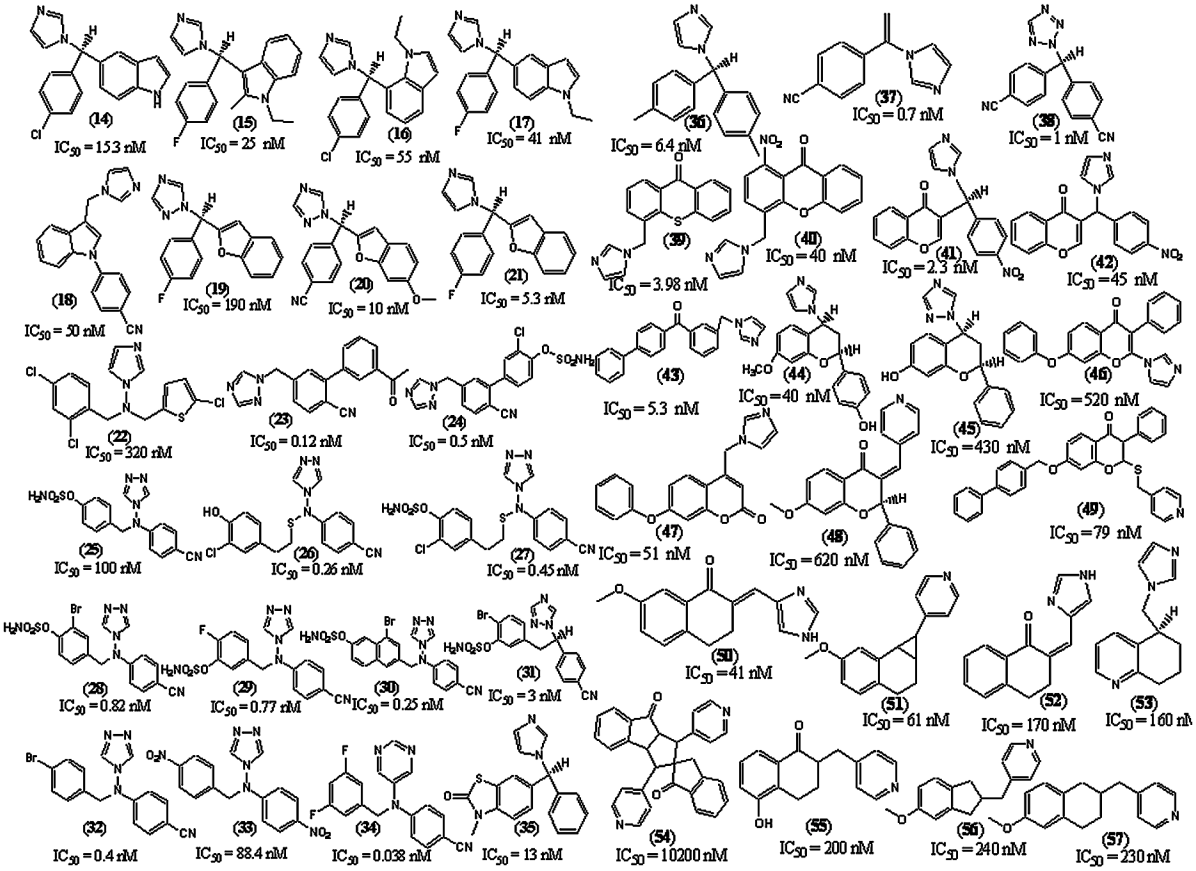
1997). It is more selective than AGT (**3**)**.** It is 700 times more potent than AGT (**3**). The FDZ (**4**) has some nonselective inhibitory activity in aldosterone, progesterone and corticosterone biosyntheses. The 'R' isomer of FDZ (**4**) is less potent than the 'S' isomer (Furet *et al*., 1993). The triazole analog ANZ (**5**) (Figure 1) is an achiral triazole derivative having potent aromatase inhibitory activity (IC_{50} = 15 nM) (Plourde, Dyroff, & Dukes, 1994). The LTZ (**6**) is also a potent aromatase inhibitor in human placental aromatase (IC_{50} = 11.5 nM) (Bhatnagar *et al*., 1990). It is found to be orally active. It causes regression of tumors in 7, 12-dimethylbenz [a] anthracene (DMBA) induced hormone dependent rat tumor model (Demers, 1994). The ANZ (**5**) and LTZ (**6**) are highly specific drugs. These are 100-3000 times more potent than AGT (**3**). It is a challenge to design higher degree of specificity and more selective aromatase inhibitors still now. Researchers are trying to design and synthesize more effective and selective imidazolylmethyl, triazolylmethyl and pyridylmethyl derivatives of NSAIs (Neves *et al*., 2009; Jackson *et al*., 2008; Furet *et al*., 1993; Leonetti *et al*., 2004; Cavalli *et al*., 2005; Le´ze´ *et al*., 2004, 2006, 2008; Marchand *et al*., 2003; Le Brogne *et al*., 1999, 2007; Vinh *et al*., 1999; Saberi *et al*., 2006; Khodarahmi *et al*., 2001; Whomsley *et al*., 1993; Woo *et al*., 2003, 2007, 2010; Bubert *et al*., 2008; Wood *et al*., 2005, 2008, 2010; Okada *et al*., 1996, 1997a, 1997b, 1997c; Serpentini *et al*., 2004; Jones *et al*., 1990; Taylor *et al*., 1987; Lang *et al*., 1993; Recanatini *et al*., 2001; Gobbi *et al*., 2006, 2007, 2010; Pouget *et al*., 2002a, 2002b, 2004; Yahiaoui *et al*., 2004; Hackett *et al*., 2005; Su *et al*., 2005; Kim, Hackett, & Bruggemeier, 2004; Hartmann *et al*., 1995, 1996; 2004; Wachter *et al*., 1996; Jacobs *et al*., 2000; Gupta *et al*., 2004; Bayer *et al*., 1991; Hartmann, Bayer, & Grun, 1994). Some of the representative molecules having potent aromatase inhibitory activity are shown in Figure 11.

LIGAND- AND STRUCTURE-BASED DRUG DESIGN (LBDD AND SBDD) OF AROMATASE INHIBITORS

The LBDD may play important roles for designing potent NSAIs. The chemometric approaches include different QSAR and pharmacophore mapping techniques. It is observed that conformational flexibility of ligands and molecular alignment represents key techniques though these are also the main difficulties in LBDD. Various molecular modeling tools involved in LBDD are SYBYL (Tripos Inc., 2012), HipHop (Accelrys Inc. 2011), Hypogen (Accelrys Inc. 2011), DISCO (Martin, 2000), GASP (Jones & Willet, 2000), GALAHAD (Tipos Inc., 2012), PHASE (Dixon *et al*., 2006) and MOE (Chemical Computing Group, 2013). The SBDD involves the macromolecular target structure or receptor-ligand complex. Structure-based

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 11. Selective potent non-steroidal aromatase inhibitors (NSAIs)



pharmacophore mapping involves the chemical feature of the active site and their spatial relationship. The tool used for this purpose is 'LigandScout' which incorporates the macromolecule-ligand complex base scheme (Wolber & Langer, 2005). Other softwares used for this purpose are 'Pocket v.2' (Chen & Lai, 2006) and 'GBPM' (Ortuso, 2006). The limitation of this approach is the need for the 3D-structure of the macromolecule-ligand complex. It implies that it cannot be applied when compounds targeting the binding site of interest are unknown. This problem can be overcome by the macromolecule-based approach. The structure-based pharmacophore method implemented in 'Discovery Studio' (Accelrys Inc. 2011) is a typical macromolecule-based approach. It converts LUDI interaction maps within the protein binding sites along with different pharmacophore features (Bohm, 1992). The main limitation of the structure-based pharmacophore flowchart is that the interaction maps generally consist of a large number of unprioritized 'Catalyst' features which complicates its application for searching 3D databases. To overcome this problem, hot-spots-guided receptor-based pharmacophores (HSPharm) is proposed (Barillari, Marcou, & Rognan, 2008). This approach may be helpful for prioritization of atoms that should be targeted for ligand binding by machine learning algorithms with atom-based fingerprints of known ligand-binding pockets. Again, an apoprotein-based approach, i.e., GRID molecular interaction fields (MIFs) (Tintori, 2008) may be helpful for analyzing binding site of interest and conversion of the minimum MIFs into pharmacophore features. However, a pharmacophore model composed of too

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

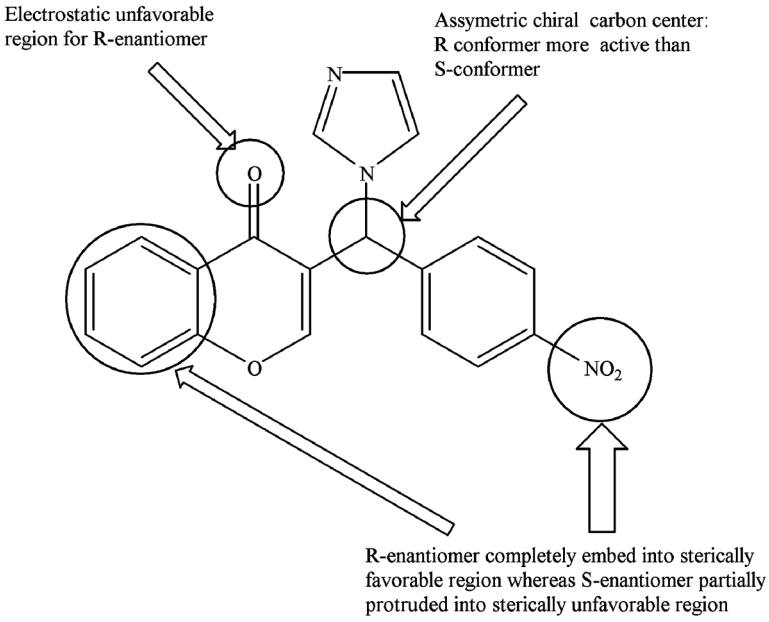
many (>7) or too less (<3) chemical features may not be suitable for 3D database screening. Hence, it is necessary to select limited number of chemical features (3 to 7) to construct a practical hypothesis. However, pharmacophore approaches are one of the most successful concepts in drug designing strategies. Combination of pharmacophore mapping followed by other molecular modeling approaches like docking may be a good strategy for drug designing approach. Compared with pharmacophore-based virtual screening, pharmacophore-based *de novo* design show unique advantage in building novel hits. The pharmacophore based design may also be extended to lead optimization (Brenk & Klebe, 2006), multitarget drug design (Wei *et al*., 2008), activity profiling (Steindl *et al*., 2006) and target identification (Rollinger *et al*., 2004).

Docking method is an *in silico* technique that determines the binding site of a protein and estimates the binding affinity of small molecule to the protein. Docking study involves the conformational and orientational degrees of freedom of a small molecule within the constraints of binding pocket of the protein. The program uses a scoring function to select the best pose of each molecule and top scoring hits may be identified. Docking method predicts the preferred orientation of ligands to a macromolecule when bound to each other to form a stable complex (Lengauer & Rarey, 1996). Molecular recognition may play the key role in promoting fundamental events like enzyme-substrate, drug-protein and drug-nucleic acid interactions. The nature of interactions like hydrogen bonding, van der Waals and electrostatic interactions may provide a conceptual framework for designing the desired potency and specificity of potential drug candidates (Mohan *et al*. 2005). Docking method includes rigid body docking (both receptor and molecule treated as rigid), flexible ligand docking (receptor treated as rigid but ligand kept as flexible) and flexible docking (both receptor and ligand kept treated as flexible). Different docking algorithms used for docking program are GOLD (genetic algorithm), AUTODOCK (Lamarckian genetic algorithm), GLIDE (Hybrid), LigandFit (shape matching based on moments of inertia), DOCK (shape matching based on sphere images), FlexX (incremental construction), Hammerhead (incremental construction), FRED (shape matching based on Gaussian functions) and Surflex (dock surface-based molecular similarity) (Kroemer, 2007). Scoring functions can be classified into three categories - knowledge-based, empirical and force field-based. Regarding the LBDD and SBDD of NSAIs, several studies were performed to design potent and specific NSAIs.

Cavalli et al. (2005) proposed a CoMFA model that represents further extension of previous ones (Recanatini & Cavalli, 1998; Recanatini, 1996) with total of 70 NSAIs. Interestingly, 7 of the newly added molecules carried at least one stereogenic center and showed enantioselectivity towards the biological target. In agreement with their previous models (Recanatini & Cavalli, 1998; Recanatini, 1996), the steric regions were located around the para-CN phenyl ring of *S*-fadrozole, whereas the para-CN phenyl ring of *R*-fadrozole partially protruded into the sterically disallowed region. Negative electrostatic blue regions were scattered around the template, whereas positive electrostatic red regions partially surrounded the negatively charged atoms of the para-CN phenyl ring of *R*-fadrozole. It indicates that the increase of positive charge may be required for higher inhibitory activity in that region, whereas the *R*-fadrazole locates its negatively charged phenyl ring. Based on that CoMFA model, some chromone analogs were synthesized keeping in mind the pharmacophoric para-substituted benzylimidazole moiety and to introduce an asymmetric center on the methylene carbon atom. The para-CN phenyl ring of the *R*-enantiomer of the best active compound was found completely embedded in the positive steric CoMFA contour. Its *S*-enantiomer placed the same moiety into a positive electrostatic region. The *R*-enantiomers were always calculated more potent than the *S*-ones. Hence, the CoMFA model can be considered as an

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 12. Structural requirements of chromones for effective NSAIs



example of enantioselective 3D-QSAR model (Cavalli *et al*., 2005). The schematic representation on the basis of the best active chromone analog for imparting aromatase inhibition is shown in Figure 12.

Some coumarins were compared with androstenedione as aromatase inhibitors by shape similarity analysis (Chen *et al*., 2004). It was found that replacement of 7-methoxy group with a hydroxyl group reduced the inhibitory activity. Again, elimination of 4′-chloro group from 3-phenyl group reduced the aromatase inhibitory activity. Hence, it is suggested that both the 7-methoxy and 3-(4′-chlorophenyl) groups are important for the anti-aromatase activity of these coumarins. In addition, 4-(4′-chlorobenzyl) group instead of a 4-benzyl group was not able to suppress aromatase activity. In the compound like 3-(4′-chlorophenyl)-7-methoxy-4-phenylcoumarin, replacement of 4-benzyl group by 4-phenyl group diminished the activity. Hence, the critical functional groups controlling the aromatase inhibition were 4-benzyl, 3-(4′-chlorophenyl) and 7-methoxy groups. The most potent compound, 4-benzyl-3-(4′-chlorophenyl)-7-methoxycoumarin ($IC_{50} = 80$ nM) was compared with androstenedione, which suggested that the coumarin ring mimic the A and B rings of the steroidal nucleus and the 3-(4′-chlorophenyl) group mimics the D ring of the androgen moiety. In the compound, 4-benzyl group aligns very closely to the C-19 methyl group of androstenedione. As C-19 methyl group pointed towards the heme group of aromatase, it is supposed that the space between the heme and C-19 was not sufficient to accommodate phenyl group, whereas benzyl group can be spatially rotate by the methylene linkage and can accommodate at the place. The 7-methoxy group was found to align closely to the C-3 keto function and 3-(4′-chlorophenyl) group aligned near the C-17 keto oxygen of androstenedione. It suggested that electron withdrawing groups might be favorable at this position. The crucial groups important for better activity are shown in Figure 13.

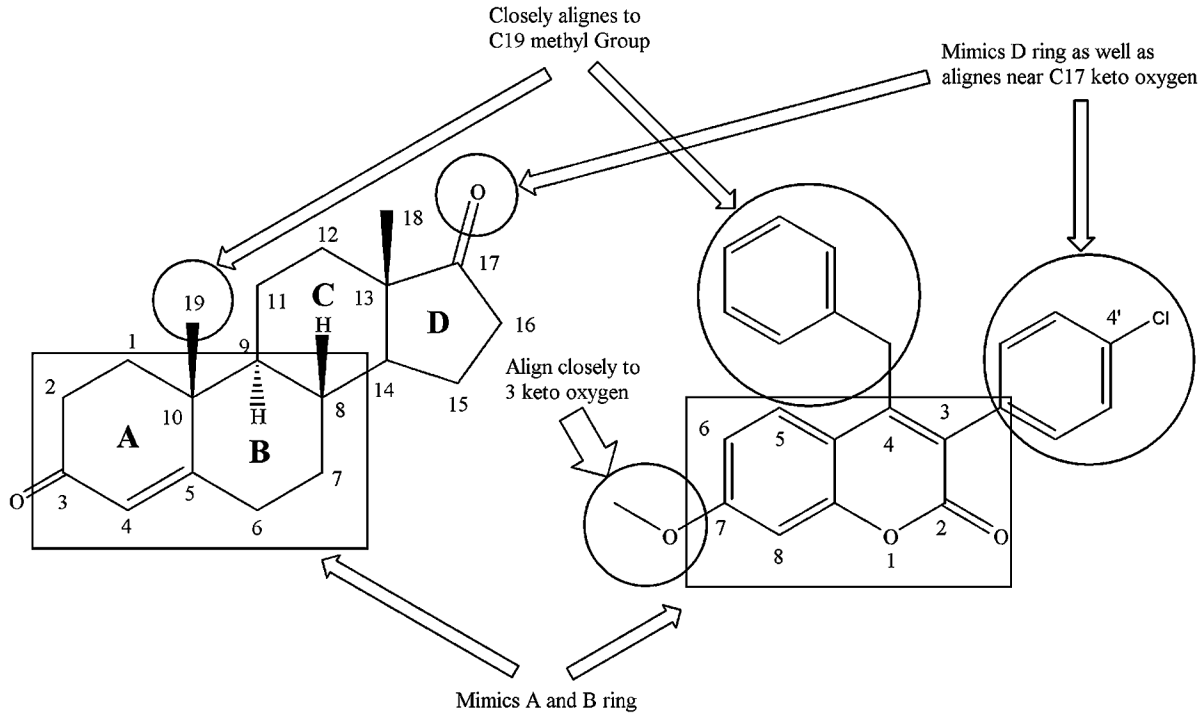
Some azole compounds linked to a fluorene, indenodiazine or coumarin scaffold were also designed and synthesized (Leonetti *et al*., 2004). Amongst these, azole-linked coumarin derivatives were found to be the most active. These coumarin derivatives were also more selective over 17-α-hydroxylase/17-20

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

lyase. CoMFA study suggested that the electronegative keto oxygen of the coumarin ring was important for the higher activity. The compounds 1, 2, 4-triazole derivatives were less active than imidazole derivatives as it was surrounded by a negative electrostatic region. Positive steric effect was important for the higher activity due to the lipophilic aromatic rings whereas the negative steric effect was not favorable for the activity. Again, methoxy group containing compounds were found to be less active. This result was in agreement with the previously published CoMFA results (Recanatini & Cavalli, 1998; Cavalli *et al*., 2000). The important structural features are shown schematically in Figure 14.

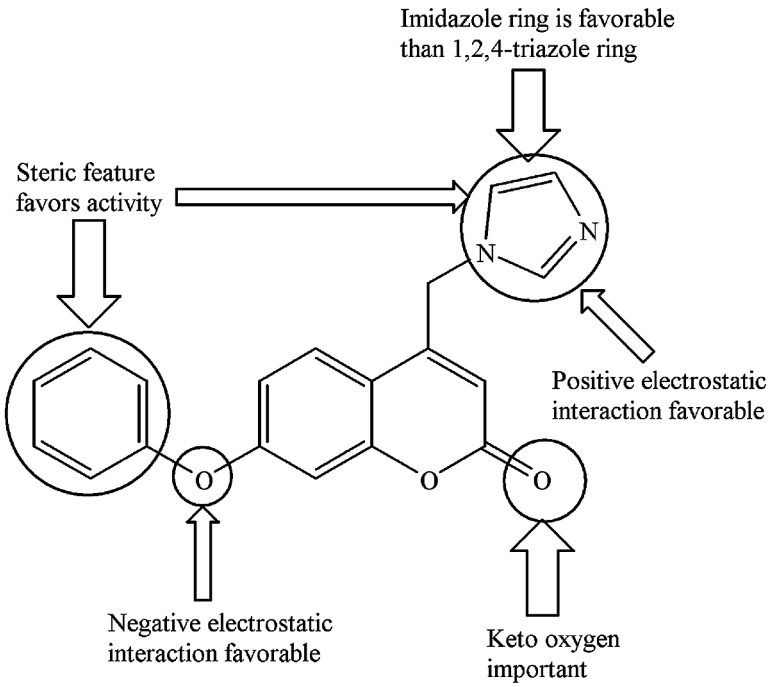
Flavone derivatives were also found to be effective aromatase inhibitors (Gobbi *et al*., 2006; Pouget *et al*., 2002). The 2D-QSAR study (Nagar *et al*., 2008) suggested that the presence of the positive charge function at the atom C_1 and decrease in the electron density at the atom C_{12} (Figure 15) favored inhibitory activity. The presence of nucleophilic substitution at meta position of the ring C, i.e., in R_2 and R_4 substitutions decreased the activity. Presence of hydrogen bond donor group at C_{15} was found to be important for the activity which was adjudged by the previous study (Cavalli *et al.* 2005). Substitution in the ring A of flavone scaffold imparted hydrophobicity of these molecules. Aromatic attachment at the atom C_8 was also found to be important for the activity. This result was also in agreement with previous findings (Recanatini *et al.*, 2001). The CoMFA study indicated favorable steric influence of imidazole ring at R′′ substituent. Additionally, substitutions in the ring C favored inhibitory activity possibly by imparting steric hindrance on that region, whereas similar substitution in the ring A (R′ position) proved to be unfavorable for the activity. The CoMSIA study suggested that the presence of acceptors at position 11 as well as in the imidazole-1-methyl ring, hydrophobic substituent at the atom C_2 and steric hindrance due to the imidazole ring and the C rings increased the inhibitory activity. Oxygen

Figure 13. Crucial structural requirements of coumarins as promising NSAIs



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 14. Important structural features of coumarin azoles required for better NSAIs

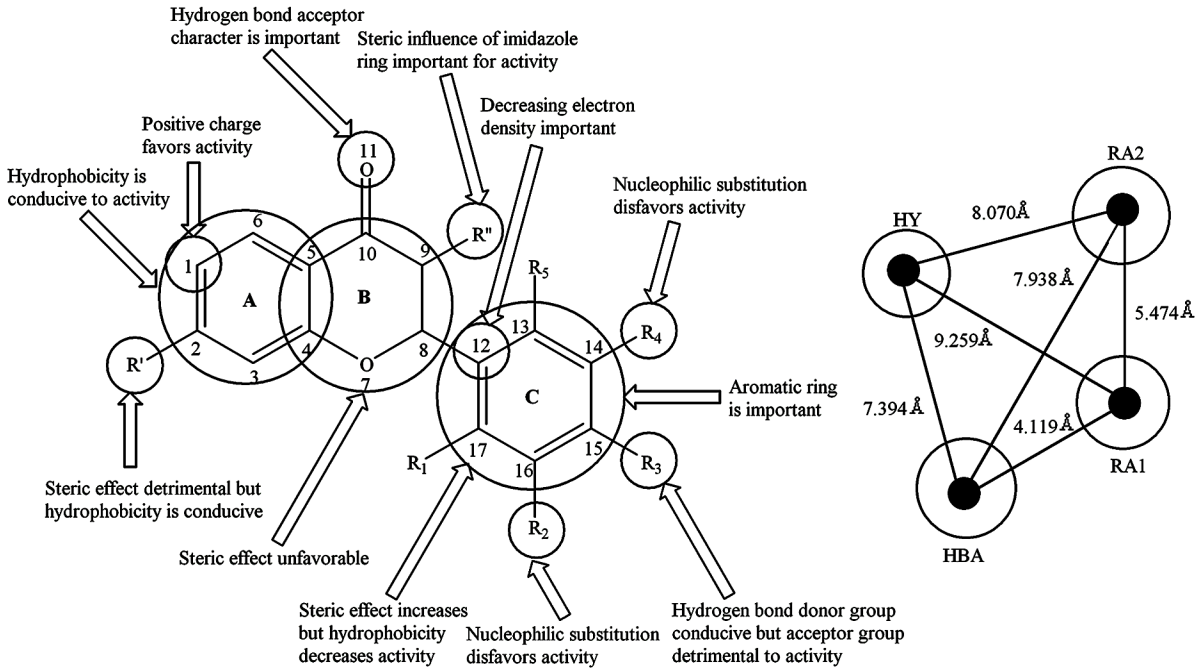


atom at position 11 might behave as a hydrogen bond acceptor that possibly binds with the Ser478 of the aromatase enzyme (Pouget *et al*., 2002). The presence of acceptors at the atom C_{15}, β-position of the imidazole ring and hydrophobic substituent in the ring C as well as steric influence at the atom C_2 and the ring B might decrease activity. The pharmacophore modeling study could also be correlated with the QSAR studies. The ring A along with its substitutional pattern imparted steric hindrance that provided hydrophobic feature of these molecules. The presence of hydrogen bond acceptor (C=O group) and the imidazole-1-methyl ring at R′′ position of the B ring increased inhibitory activity. The pharmacophore space modeling study suggested that two ring aromatic features, one each hydrogen bond acceptor and hydrophobic features are necessary for the anti-aromatase property of these molecules. The distance between the hydrophobic and the acceptor feature is 7.394 Å, whereas the two ring aromatic features are located 9.259 Å and 8.070 Å respectively away from the hydrophobic feature. Similarly, the ring aromatic features are located 7.938 Å and 4.119 Å away from the hydrogen bond acceptor feature. The aromatic rings are separated by a distance of 5.474 Å. The structural and pharmacophoric requirements of these flavones derivatives for better aromatase inhibition are represented schematically in Figure 15.

Natural phytoestrogens, such as lignans, flavones and coumestrol were found to be effective aromatase inhibitors. A molecular dynamics simulation (MDS) study was performed to understand the mechanisms involved in the binding of various phytoestrogens to the aromatase enzyme (Karkola & Wahala, 2009). The simulation study was analyzed to find the essential interactions took place on binding with the aromatase enzyme to the phytoestrogen, and were subjected to docking study to know the binding interactions. The MDS study of the aromatase-enterolactone complex suggested that there was only an occasional hydrogen bond formed between enterolactone to the backbone carbonyl of Glu302. A few water molecules were present occasionally inside the active site cavity. It suggested that the character

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 15. Structural and pharmacophoric requirements of flavones for aromatase inhibitory activity

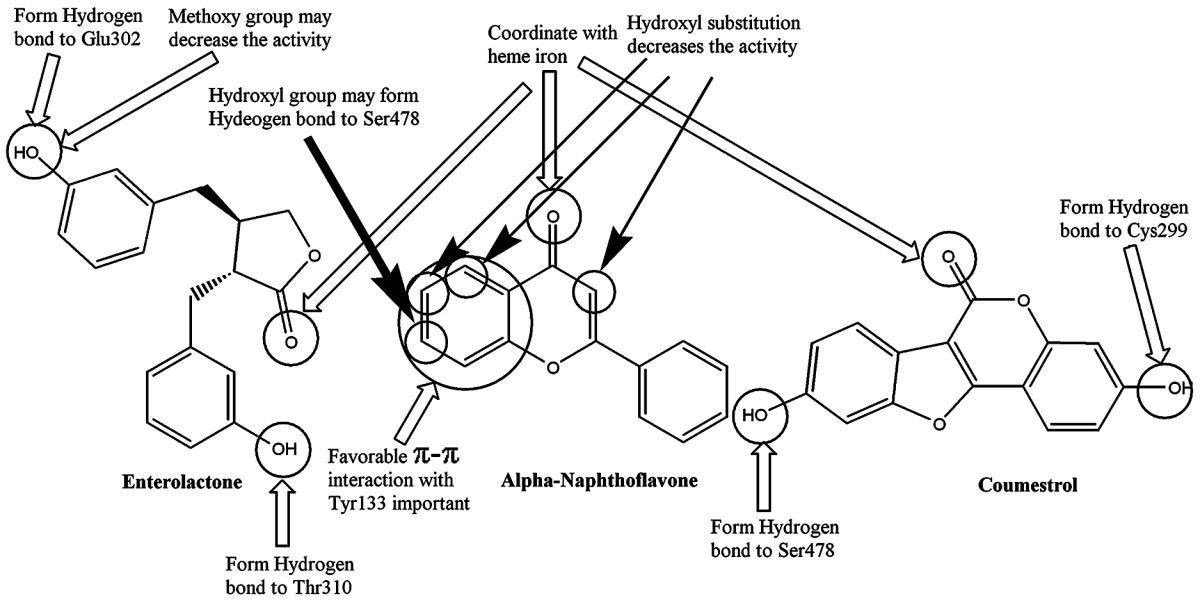


of the active site was lipophilic in nature. There was coordination between the enterolactone carbonyl oxygen and the heme moiety. The active site residues, i.e. Ile133, Phe134, Ala226, Leu227, Ile229, Lys230, Tyr244, Glu302, Ile305, Ala306, Ala307, Pro308, Asp309, Thr310, Met311, Val313, Ser314, Phe430, Cys437, Ala438, Gly439, Lys440, Leu479, His480 and Pro481 along with the heme were found in contact with enterolactone. In case of MDS of the aromatase-α-naphthoflavone complex, interaction between carbonyl function and the heme group was observed. Only a few water molecules were found to form hydrogen bonding to the active site residues, such as Glu302, Lys230 and His480. The active site residues, i.e. Leu122, Ile132, Ile133, Phe134, Phe148, Lys230, Cys299, Glu302, Met303, Ala306, Ala307, Thr310, Met374, Phe430, Cys437, Ala438, Gly439, Leu477, Ser478, Leu479 and His480 in addition to heme group were found in contact with α-naphthoflavone. The docking study of the lignans showed that compounds having lactone ring have the possibility to coordinate the heme group. Optimal coordination position and orientation would be achieved if the plane defined by the lactone group is perpendicular to the heme plane. The strength of the coordination to the heme group was also dependent on the tilting and torsion angles of the coordinating systems. Hence, a small change from the optimal coordinating geometry might lead to weaker binding. The best ranked pose showed that hydrogen bonds between the phenolic hydroxy group at the C3 atom of the ring A and the backbone carbonyl of Thr310, the hydroxy group at the C3′ atom of the ring B and the backbone carbonyl of Glu302. All compounds with a phenolic hydroxyl at the C4 atom formed a hydrogen bond to the backbone carbonyl of Pro481. The higher activity was dependent probably on the hydrogen bonding between the C3 hydroxyl group with Thr310 and a water-mediated hydrogen bond between the ring B hydroxyl group with Lys230 or Ile305. The methoxy group at the C3′ atom might block a water-mediated hydrogen bond and lead to decrease in the activity. The docking study of α-naphthoflavone showed to coordinate with the heme iron via the C4 carbonyl group. The naphthyl moiety of α-naphthoflavone was located near residues Leu122,

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Ile133, Phe134, Lys230, Val373, Met374, Ser478 and Leu479, and the phenyl B-ring near residues Ile132, Phe148 and Cys299. Hence, the hydrophobic skeleton of α-naphthoflavone was accommodated by two hydrophobic areas in the active site. A favorable π-π interaction between the aromatic tyrosine ring of 133 and the B ring of flavonoid were also observed, whereas the unfavorable π-π interaction was observed with the catechol hydroxyl group and the 7, 8-dihydroxy-flavones. It was found that Asp309 has not directly been involved with the binding of inhibitor. The hydroxyl group of Thr310 played a dual character in the active site. It might serve as a hydrogen-bonding partner of water molecules for catalytic reaction and might stabilize the link formed by Pro308 in the I-helix by hydrogen bonding to the carbonyl group of Ala306. The steric effects of the ring B attached to the C2 atom in isoflavone shifted the carbonyl group to an unfavorable position and angle compared to the carbonyl group in flavones. The coordination of isoflavone was weaker compared to flavone and thus, might lead to decrease in the activity. A hydroxy group at the C7 atom of the flavone moiety might increase the inhibitory potency significantly due to the hydrogen bonding to Ser478. Hydroxy groups at atoms C3, C5 and C6 led to decreased inhibitory activity. The hydroxyl group at the C6 atom was too far to form a hydrogen bond to Ser478, and therefore less active. A hydroxy group at the ring B also led to decreased activity. In the docking study, a hydrophobic skeleton, a sterically unhindered carbonyl functionality and a hydrogen bonding substituent at the C7 atom of the flavone skeleton might be necessary factors for effective binding. The docking of coumestrol showed that it could form hydrogen bonding to hydroxyl side chain of Ser478 and carbonyl backbone of Cys299 to the phenolic hydroxy groups. The His480 was located near the aromatic skeleton of coumestrol and the mutation to a glutamine led to the loss of π-π interactions and hence, decreased the inhibition potency. Amino acid Asp309 was not directly involved in phytoestrogen binding and therefore, the mutation to an alanine did not directly affect the binding of coumestrol. Schematic representation of the docking results for better activity is shown in Figure 16.

Figure 16. Schematic representation of the docking results of phytoestrogens for better activity



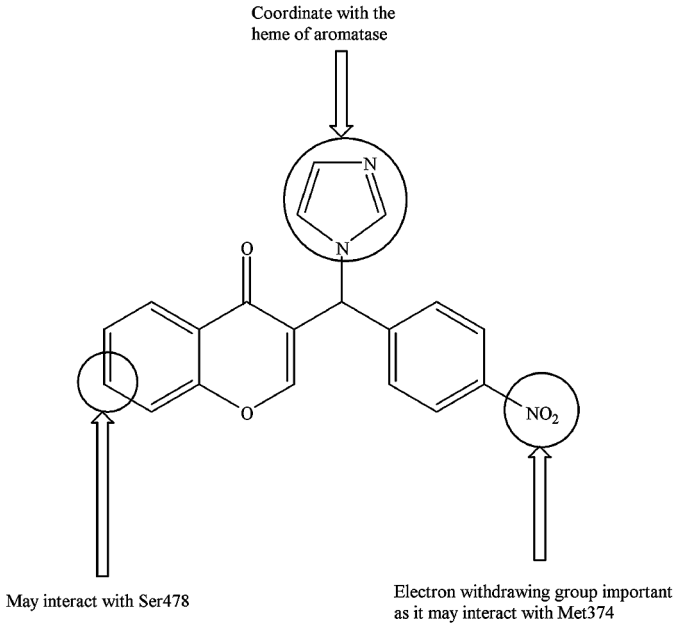
EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Flavonoid derivatives having anti-aromatase activity were subjected to docking study by calculating three types of scoring functions, i.e. knowledge-based, force field-based and empirical scores (Narayana *et al*., 2012). The best docking pose was selected based on the interactions of the imidazole nitrogen atom as well as the carbonyl oxygen atom to the heme portion of aromatase. As the individual scoring functions showed low correlation with the activity (<0.44), it was not useful to identify true hits. Thus, consensus prediction models were used for calculating aromatase inhibition. Two models were developed. One is based on the individual descriptors of each docking protocol and second on the different scoring functions of three docking methods. The descriptors obtained by the GOLD and GLIDE docking methods were correlated with the activity. Different internal and external validation parameters were used for model validation. The first model showed the importance of Astex statistical potential (ASP) and Score of internal energy [S(int)]. The ASP is the atom–atom potential derived from the frequency of interactions between ligand and receptor atoms, i.e. heme portion, whereas S(int) denotes the sum of ligand internal van der Waals energy and ligand torsional strain energy. Because of the presence of heme group, it could be inferred that the metal-ligand interaction or coordination may play a significant role of aromatase inhibition. The S(int) signifies the stability of the ligand inside the receptor active site. The second model showed the significance of the scoring functions like ASP_Score, PMF (potential of mean force) and dock score. The ASP_Score is the atom–atom potential derived from a database of protein-ligand complexes, which signifies the information about the frequency of interaction between ligand and receptor atoms. It also gives information about the statistical potential. Potential of mean force is calculated by summing up pairwise interaction terms, overall interatomic pairs of the receptor-ligand complex along with metal ion interactions and halogen potentials. Increasing value of PMF suggested that these interactions were important as far as the aromatase inhibition was concerned. The sum of ligand internal energy and the receptor-ligand interaction energy are denoted by the dock score, and it implied that both of the interaction energies were important for aromatase inhibition. To judge the accuracy of both of these models, an external validation set was used. All these external set compounds were well validated, and external validation parameters proved the reliability and robustness of the developed models. This study proved that flavonoid derivatives with imidazolylmethyl substitution were more potent than unsubstituted derivatives suggested by the docking score. Conformational and spatial arrangement of the imidazole ring toward the heme (Fe^{3+}) of the aromatase enzyme was important for higher anti-aromatase activity. In case of unsubstituted flavonoids, interaction with heme (Fe^{3+}) moiety might be observed with cyclic ketone. The methoxy group at 7-position and the electron-withdrawing groups with 4-phenyl substituent at 3-position of flavonoids were found to be more potent than the unsubstituted flavonoids. This could be attributed to H-bond interaction with Ser478 and Met374 amino acids of aromatase enzyme. These models might be useful for designing new NSAIs. Structural requirements for designing better anti-aromatase compounds of this class are shown in Figure 17.

Flavonoids are found to be effective aromatase inhibitors. The QSAR and classification structure-activity relationship (CSAR) studies were performed on some anti-aromatase flavonoid derivatives (Le Bail *et al.*, 2001; Pouget *et al*., 2002b, 2002c; Yahiaoui *et al.*, 2004; Gobbi *et al.*, 2006; Mohammed *et al*., 2011) using multiple linear regression (MLR), artificial neural network (ANN), support vector machine (SVM) and decision tree (DT) approaches. The QSAR model developed by MLR approach showed the importance of the number of rings (nCIC), number of hydrogen bond donors (nHDon), number of hydrogen bond acceptors (nHAcc), lipophilicity (ALogP), rotatable bond number (RBN) and dipole moment for regulating the anti-aromatase properties of the flavonoids. These descriptors were further analyzed through ANN and SVM approaches and found to be statistically significant. The classification

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 17. Important structural features of flavonoids for potent anti-aromatase activity



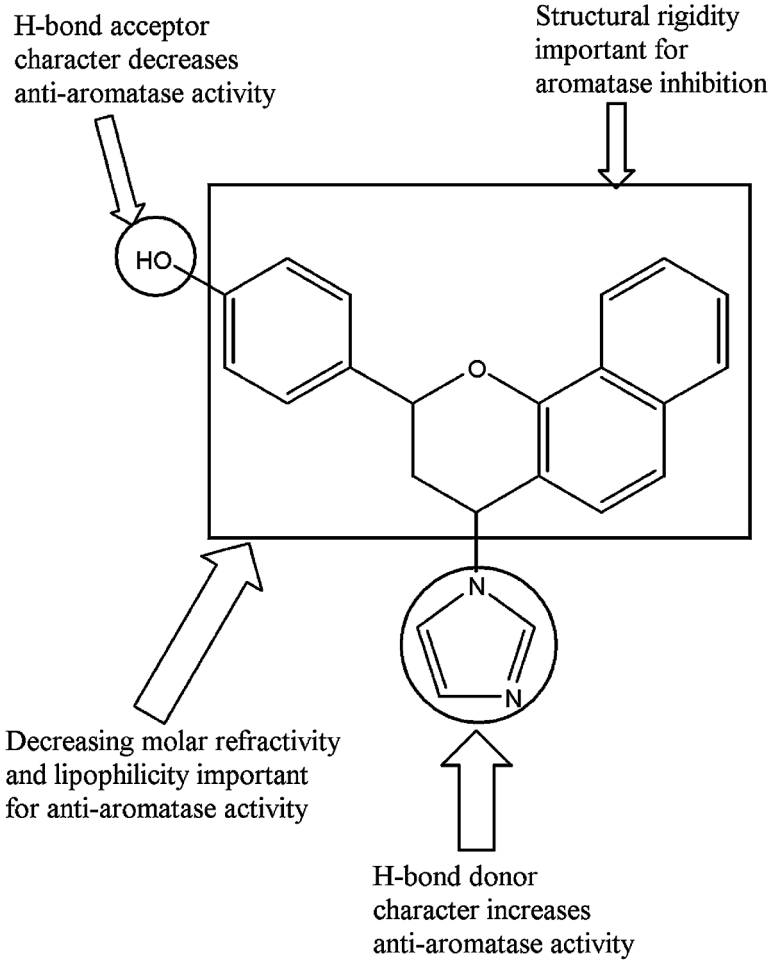
DT model proved that the nCIC, RBN and nHDon have the prime importance. This study provided that the active molecules are generally smaller, having a higher degree of rigidity, lower polarity and charge distribution, and also provided slightly lower electron-withdrawing tendency and higher chemical reactivity than the inactive molecules. The structural requirements of these flavonoids for better aromatase inhibition are represented schematically in Figure 18.

The CoMFA-based virtual screening followed by molecular dynamics (MD) simulation study was performed on a series of 45 flavonoids active against human aromatase (Awasthi *et al*., 2014). The model was found to be statistically significant and used for virtual screening of flavonoids database and followed by molecular docking study. Docking study suggested that 7-hydroxyflavanone β-D-glucopyranoside is more potent (IC_{50} = 1.09 μM) than 7-hydroxyflavanone (IC_{50} = 3.80 μM), a standardized aromatase inhibitor. This result was further validated by molecular dynamics simulation analysis and revealed the high stability and effective binding of screened flavone within the active site of aromatase.

Isoflavone derivatives were found to be effective aromatase inhibitors (Hackett *et al*., 2005; Su *et al*., 2005; Kim, Hackett, & Brueggemeier, 2004). The 2D-QSAR study suggested (Nagar & Saha, 2010c) that ionization potential, molar refractivity and electron density at atoms C_{14} and O_{18} (Figure 19) might play pivotal to the anti-aromatase activity of these isoflavones. Low electron density at the C_{14} atom and high electron density at the O_{18} atom might be favorable for aromatase inhibition. Hence, the electron withdrawing substituent at R_1 and R_2 substitutions that imparted more electron density at the O_{18} increased aromatase inhibitory activity. The CoMFA study suggested that the negative charge field at the O_{18} atom and the positive charge field at the C_{14} were important. The substitution pattern at R_3 provided more negative electrostatic force field that was favorable for the activity. The substitution pattern of R_2 imparted positive favorable steric influence on the O_{18} atom. The unfavorable steric effects are observed at atoms C_8 and C_9 positions. The CoMSIA study supported the results obtained from the CoMFA study. In addition to the steric property, the hydrophobicity of the O_{18} atom was important for

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 18. Structural requirements of flavonoids for potential anti-aromatase activity

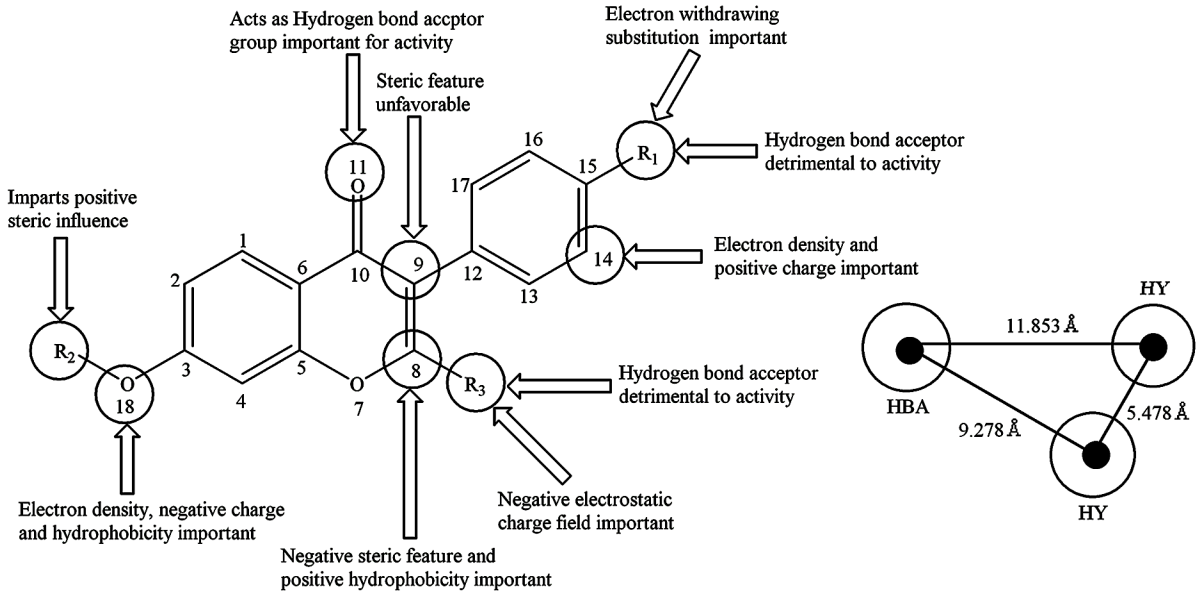


the activity. Hydrophobicity of the C_8 atom was also important for the activity. The presence of hydrogen bond acceptors at R_1 and R_3 positions had negative influence on aromatase inhibition. Pharmacophore mapping study suggested that the O_{11} atom might behave as a hydrogen bond acceptor. Hydrophobic regions around atoms O_{18} and C_8 are found to be favorable for the activity which was supported by the result of CoMSIA study. The distance between the hydrogen bond acceptor and the hydrophobic regions were found to be 11.853 Å and 9.278 Å respectively, whereas the hydrophobic regions are located 5.478 Å away from each other. Structural requirements of these isoflavones for promising anti-aromatase activity are shown in Figure 19.

Isoflavone compounds were docked into the active site of aromatase enzyme to identify the probable enzyme interactions for aromatase inhibition (Bonfield *et al.*, 2012). As the enzyme binding pocket is highly hydrophobic, alkyl chain or aromatic ring might be favorable for enhanced activity. Similarly, hydrogen bond interactions might play a significant role as the 17-keto oxygen was found to interact with Met374 backbone amide of aromatase (Ghosh *et al.*, 2009) and the hydroxyl group of Ser478 to form hydrogen bonding with carbonyl or cyano group containing aromatase inhibitors (Cavalli *et al.*, 2002; Stefanchi *et al.*, 2011). Again, Fe^{2+} in the heme group of aromatase was found to be responsible for

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

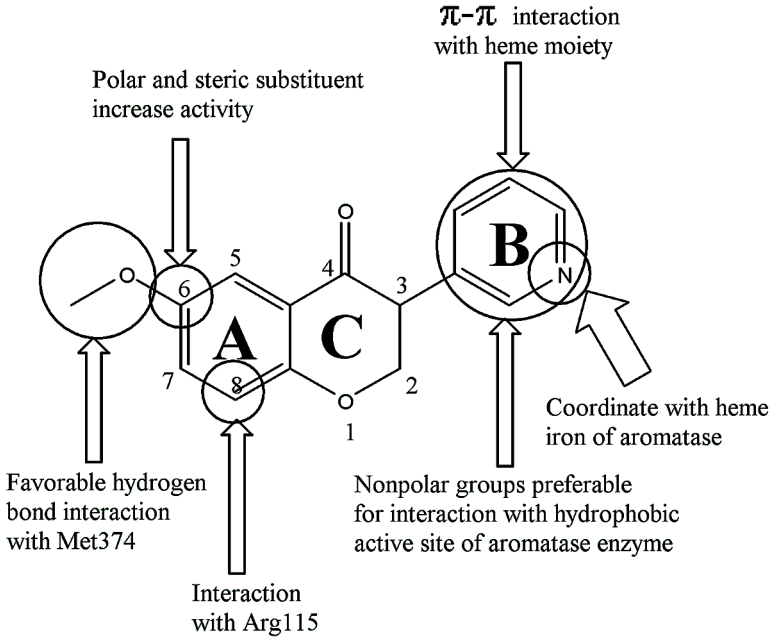
Figure 19. Structural and pharmacophoric requirements of isoflavones for anti-aromatase activity



chelating nitrogen heteroatoms of pyridine (Kim *et al.*, 2004), imidazole (Leonetti *et al.*, 2004; Yahiaoui *et al.*, 2004; Gobbi *et al.*, 2010; Recanatini *et al.,* 2001; Cavalli *et al.*, 2005) or triazole (Leonetti *et al.*, 2004; Yahiaoui *et al.*, 2004) rings. Docking study of an active compound, 6-methoxyisoflavanone (IC_{50} = 0.26 μM) showed that the methoxy group and the Met374 backbone nitrogen lied in a close proximity, indicating favorable hydrogen bonding interactions. Additionally, a π-π stacking interaction between the B ring and the heme moiety was observed, which played significant role in aromatase inhibition. But ring A was not involved in the π-π stacking interaction with the enzyme. Hence, it was inferred that increasing electron density of ring A provided no beneficial effect on aromatase inhibition though electron donating methoxy group was observed in the best active compound. Again, it was observed that relatively large polar substituent at the C_6 position might be involved in increasing the interaction between the isoflavones and the aromatase pocket. Thus, it was assumed that electron donationg substituent like methoxy and electron withdrawing substituent like benzo may increase aromatase inhibitory activity. Substitution by chlorine at C_6 position increased activity, whereas substitution by t-butyl at this position led to decrease in activity. This phenomenon might be explaind from the electronic and steric properties of the amino acid residues around C_6 position of isoflavanones. As this position is spacious and polar due to the presence of Ser478 and Arg115 lying on the pocket, a polar functional group might be favorable at this position acting through steric and electronic interactions. Additional docking study was also performed on other fluorinated and bifunctionalized isoflavanones to understand their significance in aromatase inhibition (Amato *et al.*, 2014). Docking study of the best active bifunctionalized isoflavanone (IC_{50} = 0.80 μM) showed that there was a crucial interaction between the pyridyl nitrogen of the compound and the heme iron atom (distance = 2.5 Å). It was also observed that no hydrogen bonding interaction was taking place between hydroxyl group side chain of Thr 310 and the pyridyl nitrogen of this compound. Second interaction (distance = 3.82 Å) was observed between the amino acid Arg115 and the ring A of the isoflavanone. The docking observations for promoting aromatase inhibitory activity of isoflavone are represented in Figure 20.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 20. Docking interactions of the isoflavones for obtaining potent aromatase inhibitory activity

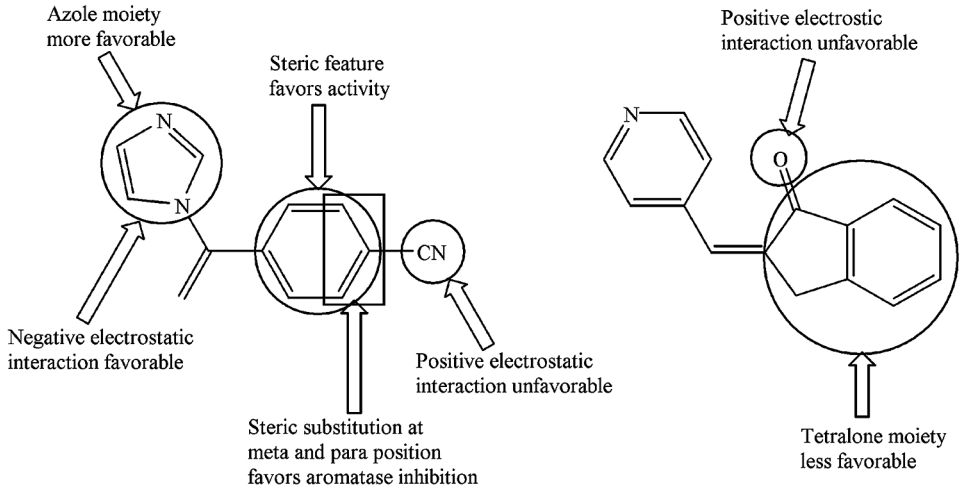


Some azoles and tetralone derivatives were subjected to CoMFA analysis (Recanatini & Cavalli, 1998). The CoMFA model suggested that the phenyl ring along with the meta and para position of azole derivatives is suitable for steric substitution. This is in agreement with previous results (Recanatini, 1996). The steric zone around the azole derivatives was found to be important for the higher aromatase inhibitory activity than tetralone derivatives, whereas steric interactions around whole molecules were found to be unfavorable for the activity. The positive electrostatic regions around the carbonyl oxygen atom of the tetralone derivatives as well as around the phenyl ring of the less active azoles were found to be unfavorable. This study also supported the previous results (Recanatini, 1996). Favorable negative electrostatic regions were observed mostly in π-electron containing the negative azole ring functions which favors aromatase inhibitory activity. Structural requirements of these azoles and tetralone derivatives for hifher activity are shown in Figure 21.

The first X-ray crystal structure of aromatase bound to androstenedione was reported in 2009 (Ghosh *et al*., 2009). In the absence of X-ray crystal structure, homology modeling is frequently performed to obtain three-dimensional structure of the desired protein. In 2000, Cavalli *et al*. reported construction of the active site of the cytochrome P450 human aromatase by homology modeling technique using cytochrome p450-CAM (PDB: 1PHD) as the template structure (Cavalli *et al*., 2000). A number of residues were found to be invariant or highly conserved in all the P450s. These include Gly60, Arg112, Gly249, Thr252, Glu287, Arg290, Arg299, Phe350, Gly353, His355, Cys357, Gly359, Ala363 and Leu375. Two structurally different non-steroidal aromatase inhibitors, i.e., S-fadrozole and a pyridyltetralone derivative were docked. Hydrophobic interactions were observed between the tetrahydropyridine moiety of S-fadrozole and the side chains of Ile305, Ala306 and Thr310. The phenyl ring of S-fadrozole may contribute to the binding of the enzyme by a number of possible hydrophobic contacts with the side chains of Val313, Thr310, Val369 and Val370. The -CN group appears to be H-bonded to the hydroxyl

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 21. Structural requirements of azoles and tetralones for promising active NSAIs



group of Ser478, and thus confirming the important role of hydrogen bond acceptor functions present in the corresponding position of fadrozole like inhibitors (Furet *et al*., 1993). This result was supported by the other model (Koymans, Moereels, & Bossche, 1995), where it was observed that cyano-substituted phenyl ring of S-fadrozole binds in a region surrounded by Asp309, Ser478 and His480. In case of the pyridyltetralone derivative, interaction was observed with pyridine nitrogen and the heme iron. A largely hydrophobic pocket made up by the side chains of Ile305, Val369, Val370 and Leu477 hosts the tetralone moiety of this inhibitor. The remarkable difference between the S-fadrazole and the pyridyltetralone is that the Ser478 side chain protrudes inside the cavity without any H-bond interaction. A CoMFA model was further developed on a related series of non-steroidal inhibitors. In case of S-fadrazole, the p-cyanophenyl ring was found to be surrounded by a favorable steric region, whereas in case of the pyridyltetralone derivatives the unfavorable steric regions at the tetralone moiety are important for aromatase inhibitory activity. Steric unfavorable region at the tetrahydropyridine moiety of S-fadrazole was also important for higher aromatase inhibition. The structural requirements of these analogs for better active NSAIs are depicted schematically in Figure 22.

Some aromatase inhibitors were synthesized and assayed followed by the 3D-QSAR study (Castellano *et al*., 2008). As per the work, analysis of the GOLPE PLS coefficients plots revealed that three areas (A, B, and C in Figure 23) surrounding the ligand-based aligned molecules were important for the activity. In all these molecules of the training set, a region was always occupied by the negatively charged substituents, thus highlighting the importance of this area in modulating the inhibitory potency. In fact, the area, i.e., ring A might be related to significant electrostatic and hydrogen bond interactions. On the other hand, the poor importance of large polyhedrons in the C region might suggest that this region was not significant for the activity. Nevertheless, most of the training set molecules were highly superimposed in this area leading to low standard deviation on the grid points surrounding it. As a matter of fact, the PLS algorithm considered the regions corresponding to the missing polyhedrons as poorly correlating with the training set activities. Again, the azole function of the B position might be responsible for coordinating with the heme iron of aromatase, which has a positive impact on aromatase inhibition. The structural requirements for better aromatase inhibitory activity are shown in Figure 23.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 22. Structural requirements of S-fadrozole and pyridyltetralone derivative for NSAIs

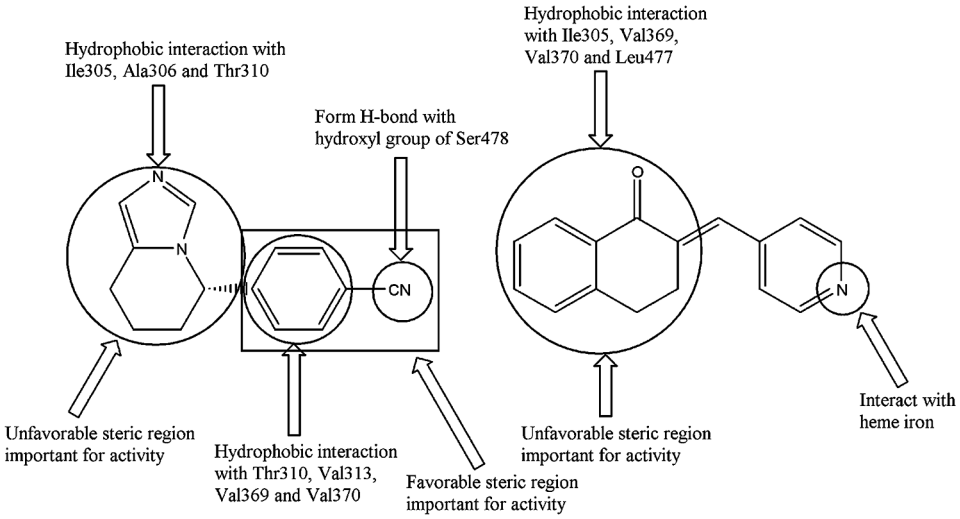
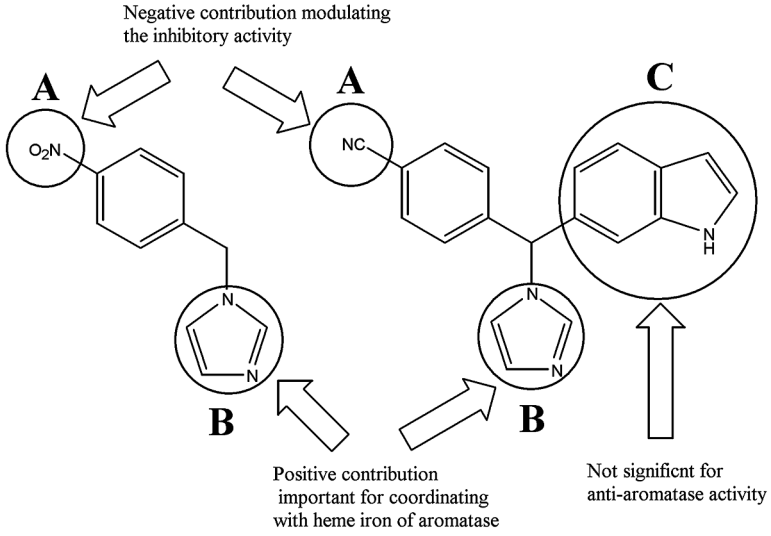


Figure 23. Important structural features required for better NSAIs



Diarylalkyl-imidazole and -triazole derivatives were found to be promising aromatase inhibitors (Karjalainen *et al.*, 2000). The QSAR model (Nagar & Saha, 2010b) revealed the importance of van der Waals atomic and molecular surface areas, presence of hydrogen atoms, importance of size and potential energy for the anti-aromatase activity of these compounds. The CoMFA study showed that regions near the imidazole or the triazole ring, and the ring A as well as substitution in the diarylalkyl chain imparted steric influence which favors the activity. The presence of the positive charges around the imidazole or the triazole ring and the negative charges around the two phenyl rings favored the activity. Hydrophobicity and electrophilic substitutions at the C_4 atom were found to be favorable for the activity. The CoMSIA

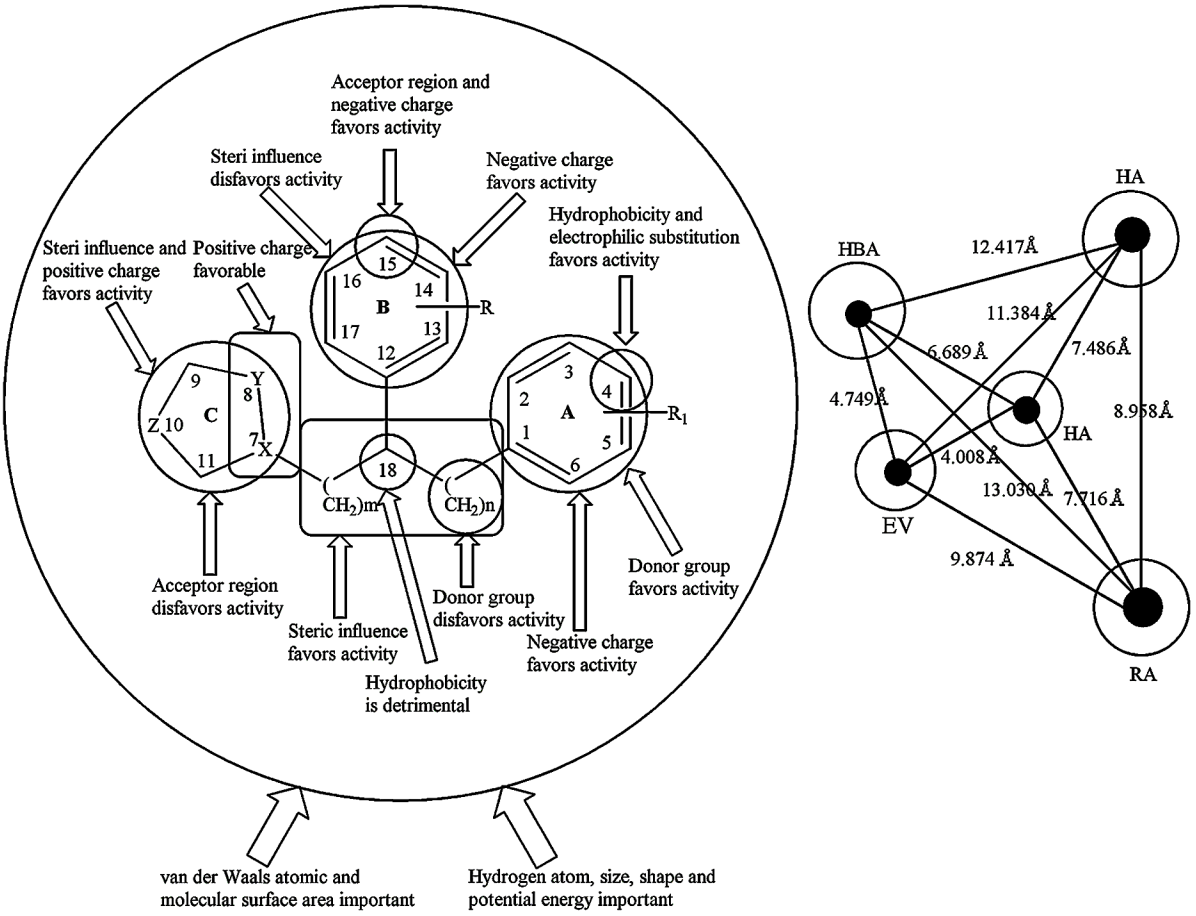
EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

study suggested that the presence of acceptor region near the ring C had negative impact, whereas the presence of the same near the ring B at C_{15} position was favorable for the activity. The steric fields around the ring C and the long aryl chain attached to the ring A, area near atoms C_{16} and C_{17} of the ring B had positive effects. Other regions of ring B and further substitution on that long chain imparted negative steric influence towards anti-aromatase activity. Favorable hydrophobic region is observed around the ring A, whereas the hydrophobic area around the C_{18} atom was unfavorable for activity. The presence of the positive charge at atoms C_{14}, C_7 and C_8 of the ring C and negative charge substitution at the atom C_{15} favored the activity. The presence of donor groups in the vicinity of the ring A was found to be favorable but donor groups at substituted long alkyl chain attached to ring A disfavored the activity. The results were adjudged through receptor-dependent modeling studies (Karkola & Wahala, 2009; Saberi *et al*., 2006), where it was observed that molecular hydrophobicity and hydrogen bond interactions played pivotal role in aromatase inhibition. The pharmacophore mapping of these molecules showed the importance of one hydrogen bond acceptor (HBA), one ring aromatic (RA) and two hydrophobic aromatic (HA) features. In hyporefine hypothesis, one excluded volume (EV) was observed to be another steric feature important for anti-aromatase activity. In the hyporefine model, the distance between the acceptor feature and the ring aromatic feature was 13.030 Å, whereas both of the hydrophobic aromatic features were located 12.417 Å and 6.689 Å away from the acceptor feature, respectively. Both of the hydrophobic aromatic features were located 7.176 Å and 8.958 Å away from the ring aromatic feature respectively, whereas the hydrophobic aromatic features were separated by a distance of 7.486 Å. The study revealed that high electronegative hydrogen bond acceptor group at the C_{15} atom, aromatic rings like imidazole or triaole with hydrophobic character and steric interaction in the ring B are crucial for the activity. Structural requirements for potent aromatase inhibitors of diarylalkyl-imidazole and -triazole derivatives are shown in Figure 24.

A diverse set of NSAIs were subjected to the docking and 3D-QSAR studies (Roy & Roy, 2010b). The docking study suggested that the ligands bind with both polar, i.e. Arg115, Arg375, Asp309, Asp371, Ser478, Thr310, Asp371, Glu302 and non polar, i.e. Ala306, Ala307, Ile133, Ile305, Leu477, Met374, Phe134, Phe221, Trp224, Val369, Val370, Val373 amino acids. This study was in an agreement with the crystallographic data that showed androstenedione-bound human placental aromatase formed a hydrogen bond with the amide of Met374 (Ghosh *et al*., 2009). The docking results indicated the important interactions observed in the active site cavity of aromatase are Met374, Arg115, Ile133, Ala306, Thr310, Asp309, Val370 and Ser478. All these compounds are found to form hydrogen bond with the amide backbone of Met374. Moreover, the nitro group containing compounds are found to form hydrogen bond with Arg115, whereas the cyano group containing compounds formed hydrogen bond with the hydroxyl group of Thr310. Apart from forming hydrogen bonds, van der Waals interactions were observed with different amino acids like Ala306, Thr310, Trp224, Val370, Ile133, Phe134, Leu372 and Val373. The azole moiety was found to be responsible for coordinating with the heme iron. Steric bumps of these compounds with different amino acids, like Asp309, Thr310, Met374, Arg115, Ser478 and Val370 might be responsible for poor activity of these compounds. The QSAR study suggested that the optimum number of hydrogen bond acceptor should be less than or equal to 2 favored aromatase binding. This result supported the docking results also. The QSAR study also suggested the importance of different shape, structural, topological and electrotopological state index that were crucial for aromatase inhibition. Optimum hydrophobicity (aLogP < 4.7) of the ligand might play vital role for aromatase inhibition. The validated QSAR study also suggested that the cyano group containing compounds might

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 24. Structural and pharmacophoric requirement of dialkyl azoles for promising anti-aromatase activity

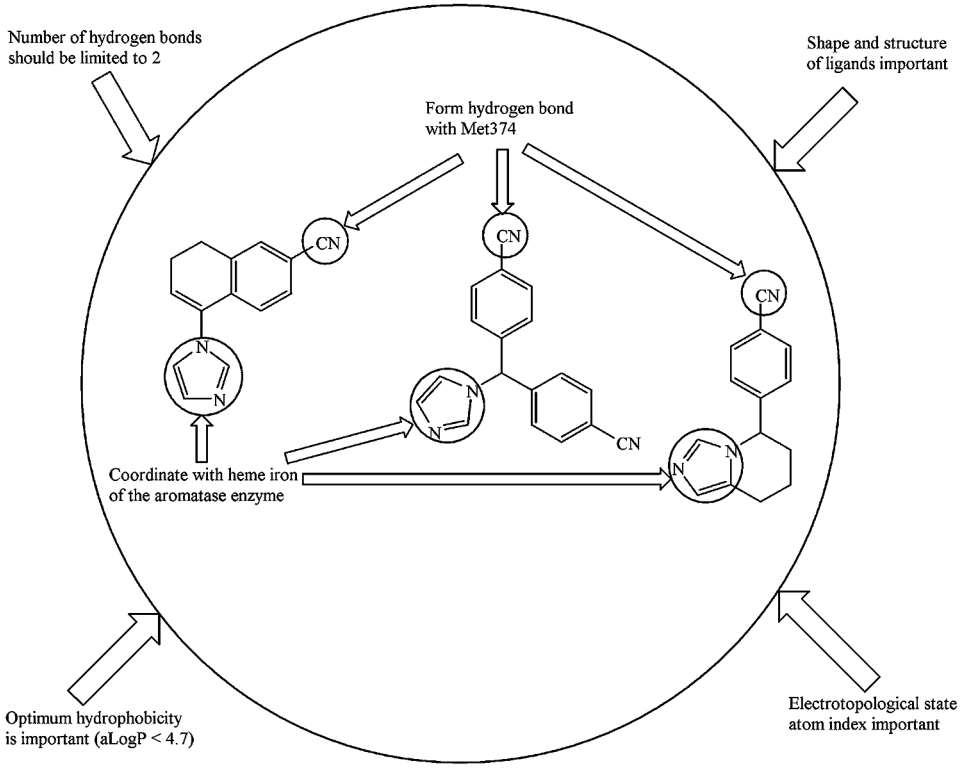


form hydrogen bond with Met374 which is supported by the previously published result (Castellano *et al*., 2008). Structural requirements for better anti-aromatase activity are shown in Figure 25.

A diverse set of molecules comprising NSAIs and SAIs were subjected to different univariate (statistical and histogram analysis) and multivariate (decision tree and principal component analysis) approaches by using quantum chemical and molecular descriptors to understand the origin of anti-aromatase properties of these molecules (Nantasenamat *et al*., 2013a). Fragment based substructure analysis provided important information about the fragment important to be an active as well as inactive for aromatase inhibition. It is observed that five of the top ten active fragments contain azole ring function, amongst them the top ranked fragment contains 1, 3 imidazole ring system as per molecular score. Apart from that, 1, 2, 4-triazole ring system was found to be an active fragment. On the other hand, 4-[(4-cyanophenyl) methyl] benzonitrile, a substructure of third-generation NSAI, letrozole (**6**), was found to be an important active fragment. Interestingly, it was observed that all the top ten inactive fragment were composed of 5-hydroxy-2-phenyl-4H,10H-pyrano[2,3-f]chromen-4-one, fusion of flavones and 4H-pyran ring. Comparing the mean activity and greater percentage of molecular fragments occurring in the active set, it was assumed that the NSAIs are more effective that SAIs. This chemical space analysis

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 25. Structural requirements for anti-aromatase activity of non steroidal compounds

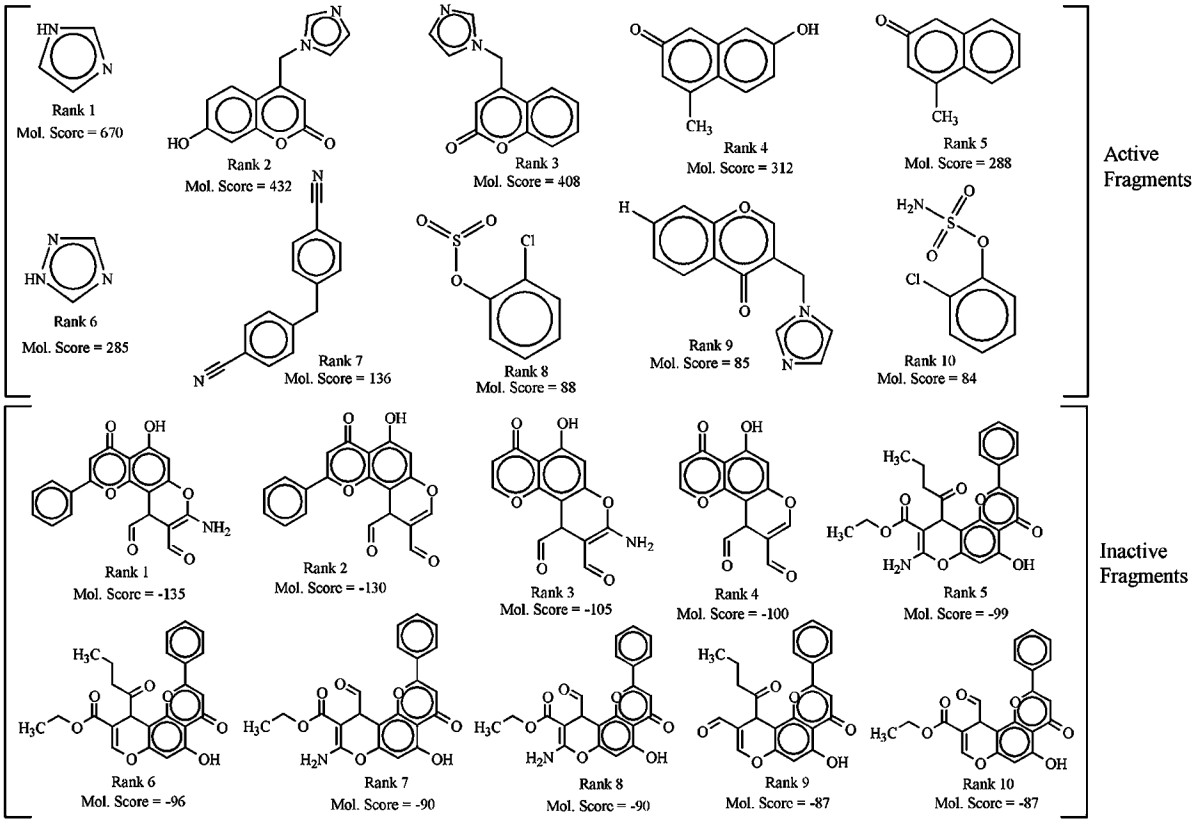


might be beneficial for future designing of potential active NSAIs. The structures of the top ten active and inactive fragments according to the molecular score as per the study are shown in Figure 26.

Large scale QSAR study was performed on the same structurally diverse aromatase inhibitors (SAIs and NSAIs), described in last paragraph, based on Monte Carlo approach (Worachartcheewan *et al.*, 2014). Such type of QSAR models might be very useful as far as the structural point of view, as this approach directly dealt with the relation between the biological activity and the SMILES notation encoding molecular structure. In this approach, the data was divided into four types of random split sets, i.e., sub-training, calibration, test, and validation sets. QSAR models were judged by parameters, like R^2 and Q^2 which correspond to the goodness-of-fit and the goodness-of-prediction parameters, respectively. Fischer (F) ratio also provided further estimate of the predictive ability of the model. The statistical results obtained from different splits and data sub-sets indicated that the QSAR model was reliable and significant. The SMILES molecular fragments might also be useful for prediction and interpretation of the origins of aromatase inhibitory activities. The top ranking fragments for imparting anti-aromatase activity include the presence of the cyclic ring in the molecular structure, absence of halogens, presence of double bond and oxygen atoms together with double bonds those are not connected in the structure. Similarly, the top ranking fragments for reducing anti-aromatase activity were branching in the molecule, presence of nitrogen and double bonds those are not connected in the structure and presence of oxygen atoms connected through double bonds. This Monte Carlo type of QSAR technique might also be beneficial for providing the mechanistic information of the molecular fragments for influencing aromatase inhibitory activity.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 26. Active and inactive fragments with their molecular score important for aromatase inhibition

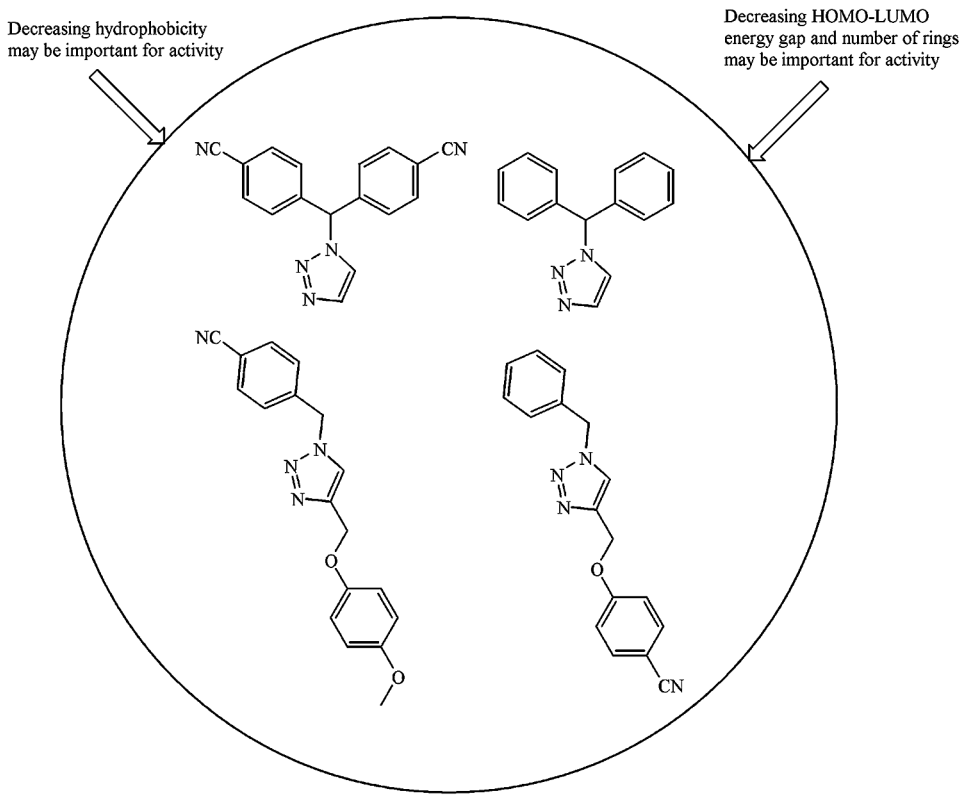


A QSAR study on 1-substituted 1, 2, 3- triazole analogs of letrozole was done (Nantasenamat *et al*., 2013b) using multiple linear regression (MLR), artificial neural network (ANN) and support vector machine (SVM). Geometry optimizations of these compounds were done by AM1 method followed by DFT method. Quantum chemical and molecular descriptors were selected for MLR model development. After removal of the outliers, the best model was subjected to ANN and SVM techniques for parameter optimization. This model might be useful to predict new aromatase inhibitors of this class. Among all the multivariate analyses performed, SVM was found to be more superior to MLR and ANN. The SVM provided the highest accuracy, whereas the MLR model could be easily interpretable. Three descriptors, i.e., number of rings, hydrophobicity and HOMO-LUMO energy gap were found to be important for aromatase inhibition. Decreasing values of these three parameters may be beneficial for the aromatase inhibition. Structural requirements for higher aromatase inhibitory activity of this class of compounds are shown in Figure 27.

A mechanism-based SAR categorization framework highlighting the important chemical and structural features responsible for aromatase inhibition was developed (Petkov *et al.*, 2009). The steroidal scaffold is found to be the most prominent when the structure of the target chemical is similar to androstenedione and testosterone. The NSAIs should possess heteroatoms like N, O and S usually in a diazole or triazole ring so that these can coordinate the heme iron of aromatase enzyme and interfere with the steroid hydroxylation. Chemical groups like CN, NO_2, CO, etc. of the NSAIs were able to accept a hydrogen bond

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 27. Important structural features of 1-substituted 1, 2, 3- triazole analogs for anti-aromatase activity

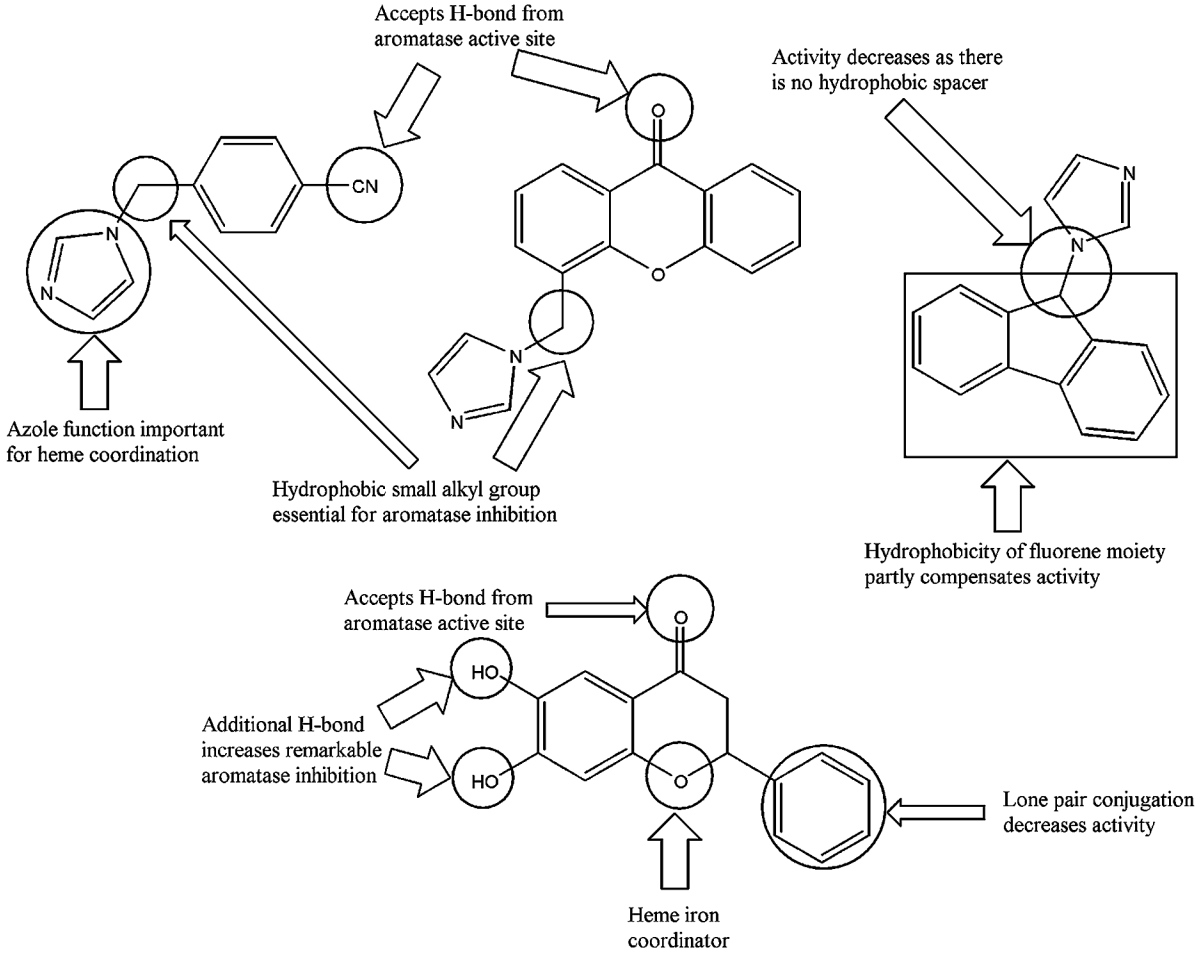


from the aromatase active site due to their electronegativities. A hydrophobic spacer group between the heme coordinating group and the hydrogen bond acceptor moiety was essential for the aromatase inhibition. However, larger alkyl fragments as spacers disfavored potency due to steric hindrance. For classes like quinolines, tetralones, flavones, etc, where no Csp3 spacer was available between electron donating and accepting centers, the activity decreased. However, in case of fluorenes hydrophobicity might partly compensate the lack of electron accepting groups. The structural requirements for better activities of these NSAIs are shown in Figure 28.

A new ligand-based strategy combining important pharmacophoric and structural features were developed according to the postulated aromatase binding mode that may be useful for virtual screening of new potent NSAIs (Neves *et al.*, 2009). Some known NSAIs were used to derive and validate pharmacophore model for fast high-throughput *in silico* screening of anti-aromatase compounds. Common feature pharmacophore model was first developed with fadrozole (**4**), anastrozole (**5**), letrozole (**6**) and vorozole (**7**) by using HipHop algorithm. Six possible alignment solutions containing three or four pharmacophoric points were found. The top two ranked solutions contained four pharmacophore features (two hydrophobic and two hydrogen bond acceptor groups). The hypotheses revealed that the imidazole or the triazole nitrogen atom acted as a hydrogen bond acceptor and might coordinate with the heme iron of aromatase active site. The second hydrogen bond acceptor feature matches either a cyano group (anastrozole, letrozole, and fadrozole) or the nitrogen atom of a methylbenzotriazole (vorozole). One of the hydrophobic features superimposed with the phenyl rings of all these molecules, and second

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 28. Structural requirements of NSAIs for better activity



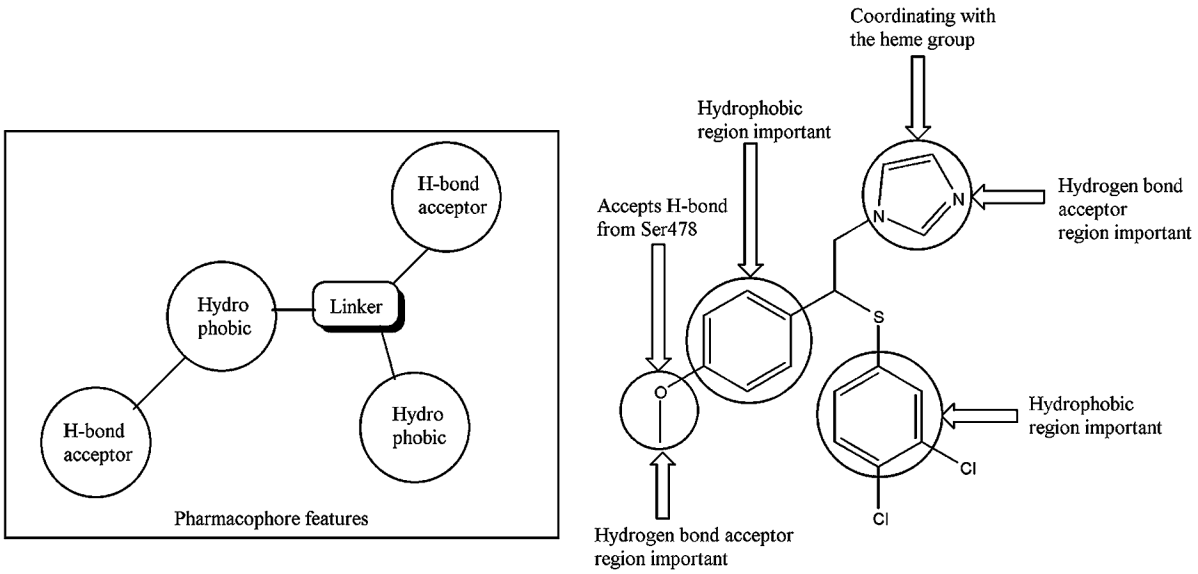
hydrophobic feature depicted the methyl group of anastrozole or the piperidine moiety of fadrozole. Besides the physicochemical properties of the inhibitor, molecular shape was found to be extremely important. Most aromatase inhibitors used to build the common features model had a similar shape that is complementary to the volume of the aromatase active site. Therefore, the shape of letrozole (**6**) is converted into a set of inclusion volumes and combined with the pharmacophoric query. Hence, the final pharmacophore model combined pharmacophoric information (two hydrogen bond acceptors and two hydrophobic groups) with fragment information (nitrogen containing aromatic heterocycles) and steric restrictions. All aromatase inhibitors used in the training set matched the final hypothesis. These findings showed that the HBA-shape pharmacophore model was able to identify not only inhibitors from the training set but also potent NSAIs with different structures. By means of virtual screening and Lipinski filtering from National Cancer Institute (NCI) database, 29 compounds were selected as hits. Four different types of nitrogen-containing aromatic heterocycles, namely imidazoles, triazoles, pyridines and pyridazines were retrieved as hits. These nitrogen containing heterocycles are linked to several types of hydrophobic molecular scaffolds, such as diphenylmethane, arylbenzylsulfane, diphenylimidazole, pyridinylbenzoimidazole, phenylpyridine, benzyl-1*H*-indol, acridine, benzo/methylfuran, quinoline, etc.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Biochemical evaluation of these compounds was performed and one compound was found to be more active than letrozole (**6**), whereas five other compounds were more active than formestane (**1**). Docking study suggested that all these molecules fit well within the binding site cavity of the aromatase enzyme, and pyridine or imidazole moiety was found to coordinate with the heme group. All molecules have a hydrogen bond between an acceptor group and Ser478 which was a key residue to the catalysis mechanism of the enzyme. Another hydrogen bond was observed between hydroxyl group and Thr310. The hydrophobic area was depicted by three phenyl rings docked on an apolar region of the aromatase binding cavity, defined by hydrophobic residues Val 313 and Leu 477. Additional hydrophobic area had either phenyl rings or a benzylsulfanyl group fitting a small hydrophobic pocket due to Pro368 and Val370. The result is represented schematically in Figure 29. This study provided useful information about the common pharmacophore features of azole NSAIs, binding mode to the active site, steric restrictions and a drug likeness filter.

Some resveratrol analogs were designed, synthesized and their inhibitory activities against aromatase enzyme were evaluated (Sun *et al*., 2010). These compounds were subjected to docking, molecular mechanics energy minimization, computer graphics molecular modeling studies, and these informations were utilized to design several very potent inhibitors. The aromatase inhibitory activity of these newly designed compounds was far better than resveratrol. One of these compounds was docked with the aromatase enzyme. The docking result suggested that the amino group of that compound may form hydrogen bonding with carbonyl group of Ile305 and carboxyl group of Asp309. The ring A might be involved with the π-π interaction with the indole ring of Trp224 and a CH-π interaction with the methyl group of Thr310. The methoxy group of the ring B might form hydrogen bond with the backbone amino group of Met374 as well as the ring B was involved with methyl group of Val370 through CH-π interaction. This docking result is in agreement with the crystallographic data (Ghosh *et al*., 2009). These docking results were used for designing new analogs. The amino group of the ring A was replaced with hydrogen bond donating groups like hydroxy and aminomethyl groups that might have similar interactions with

Figure 29. Schematic representation of the pharmacophore requirements for potent NSAIs

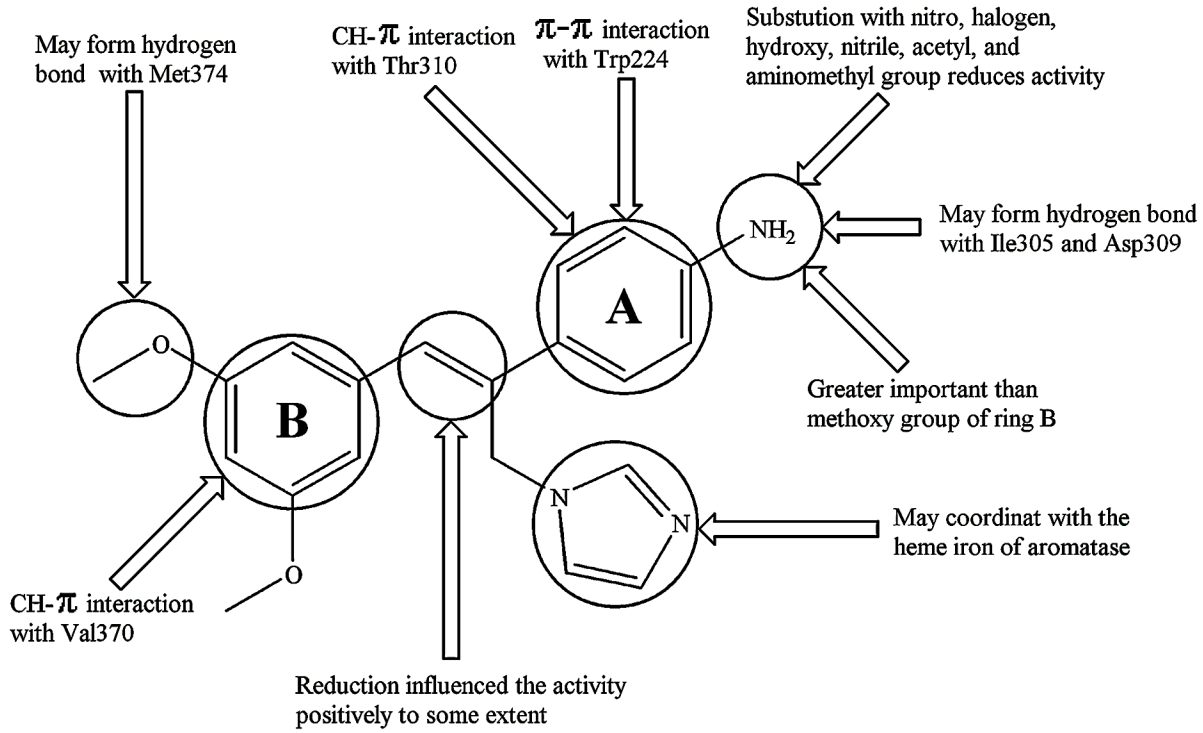


EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Asp309 and Ile305. A hydrogen bond acceptor located near the ring B may form hydrogen bond with the carbonyl group of residue Leu477. As many of these aromatase inhibitors possess imidazole and triazole groups that could coordinate to the iron atom, the azole function was incorporated into stelbene framework. Most of the resveratrol analogs with an amino group on the para position of the ring A exhibited significant aromatase inhibition. Substitution of para amino group of the ring A with nitro, halogen, hydroxy, nitrile, acetyl and aminomethyl groups resulted inactive compounds. Changing of the para amino position also reduced the activity. Reduction of the CH=CH double bond of resveratrol analog influenced the favorable activity to some extent. The para amino group of the ring A had greater importance than the methoxy group of the ring B. Incorporating imidazole group played important role by coordinating with heme iron. Structural requirements of these resveratrol analogs for higher anti-aromatase activity are shown in Figure 30.

The QSAR studies for exploring structural requirements of the dichlorophenyl compounds (Nagy, Tokarski, & Hopfinger, 1994) for better active NSAIs are represented schematically in Figure 31. The distance (D) of the nitrogen atom in the variable heterocycle from the core carbon atom was the most important factor for higher activity. To maximize inhibitor potency, the optimum distance between the nitrogen and carbon, was about 3.6 Å for these analogs. The field calculations coupled with the QSAR studies suggested that this nitrogen (located from 3.6 Å distance from the core carbon) acted as a hydrogen bond acceptor. Two possible three-dimensional pharmacophores were proposed for effective aromatase inhibitors. All these QSAR models developed showed that the distance (D) between the methane core carbon and the heterocyclic nitrogen was the crucial factor for aromatase inhibition. The steric bulk of the substituents about the chiral carbon center markedly limited conformational flexibility so that the

Figure 30. Structural requirements of resveratrol analogs for anti-aromatase activity

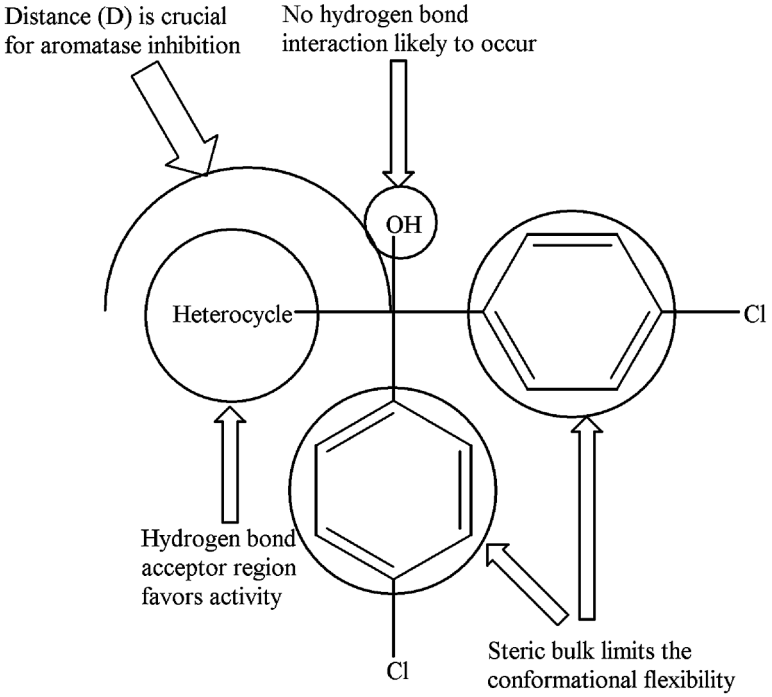


EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

global minimum energy conformation was energetically the only plausible conformation which could be considered an 'active' conformation. The distance (D) and the integrated potential field difference (ΔF) of specific region in space correlated with aromatase inhibitory activity. It suggested that hydrogen bonding or strong electrostatic interaction force between the nitrogen in the inhibitor heterocycle and the receptor was important for the higher activity. Conformational and field analyses data also indirectly supported that dipole properties of the heterocycle did not correlate with the aromatase inhibitory activity. Pharmacophore mapping study suggested that hydrogen bond acceptor regions were located at the heteroaromatic ring and in the hydroxyl group. No intermolecular hydrogen bonding was likely to occur as the site was within the receptor pocket.

A ligand-based approach was used to generate pharmacophore models derived from highly active and selective P450 19 inhibitors (Schuster *et al*., 2006). In the common feature pharmacophore model generation, hydrophobic aromatic (HAR) and ring aromatic (RA) features were considered as most frequent for active aromatase inhibitors. Hiphop pharmacophore model was generated with two ring aromatic and two HBA+F+Cl features. A Hyporefine pharmacophore model was also derived with these compounds with *in vitro* activities ranging from 40 nM to >250 μM. The HypoRefine hypothesis was generated with one RA, one hydrophobic, two HBA+F+Cl features, and one excluded volume. The Hyporefine model was able to predict 90% of all tested compounds. Using the Hiphop model as search queries and employing a sophisticated filtering procedure, the drug likeness of the test compounds were increased and two compounds were discovered. The models were used as database search queries to identify active compounds from the NCI database. The pharmacophoric requirements for better active compounds are shown in Figure 32.

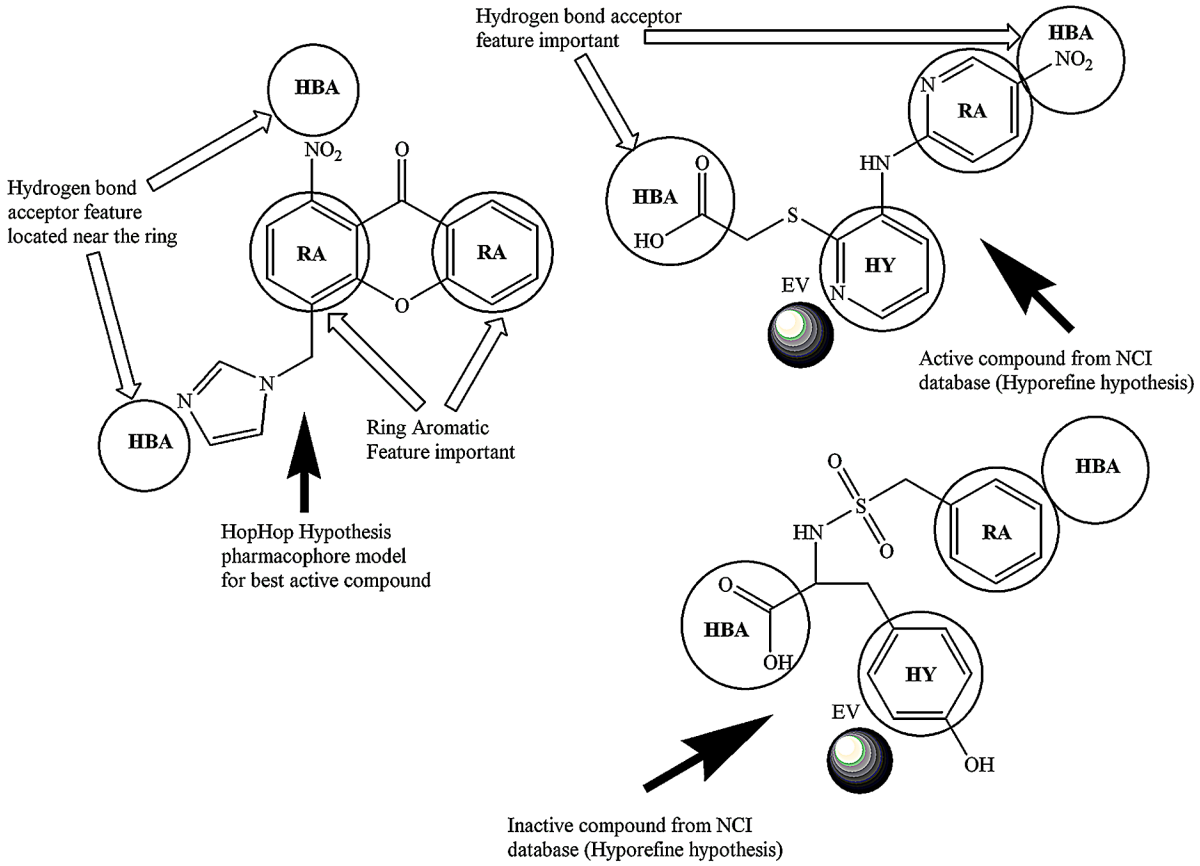
Figure 31. Structural requirements of the dichlorophenyl compounds for NSAIs



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Ligand based (LB) and structure based (SB) models were generated on a diverse set of NSAIs (Muftuoglu & Mustata, 2010) and both of these models were merged to get a new model that produced a combined pharmacophoric requirements to differentiate active and inactive NSAIs. The LB pharmacophore model consisted two hydrogen bond acceptors (A) and two hydrophobic aromatic (HYARO) features, whereas the SB pharmacophore included one acceptor (A) and two hydrophobic aromatic (HYARO) features including eleven excluded volumes (EV). In case of LB pharmacophore model, the hydrogen bond acceptor features were separated by a distance of 2.11 Å, whereas the HYARO features were separated by a distance of 4.46 Å. The distance between one acceptor and the HYARO features were 1.14 Å and 5.40 Å respectively, whereas the distances between second acceptor with the HYARO features were 3.24 Å and 7.20 Å, respectively. In case of SB pharmacophore model, the HYARO features were located from the acceptor feature at the distance of 3.16 Å and 6.99 Å respectively, whereas the HYARO features were separated by a distance of 4.46 Å. Amalgamated pharmacophore model consisted of two acceptors (A), two hydrophobic aromatic (HYARO) features and eleven excluded volumes (EV). In this model, the hydrogen bond acceptors were separated by a distance of 2.06 Å, whereas the HYARO features were separated by a distance of 4.46 Å. The distances between one acceptor and the HYARO features were 1.14 Å and 5.36 Å respectively, whereas the distances between second acceptor with the HYARO features were 3.19 Å and 7.10 Å, respectively. All these three models successfully identified

Figure 32. Important pharmacophoric features required for better NSAIs

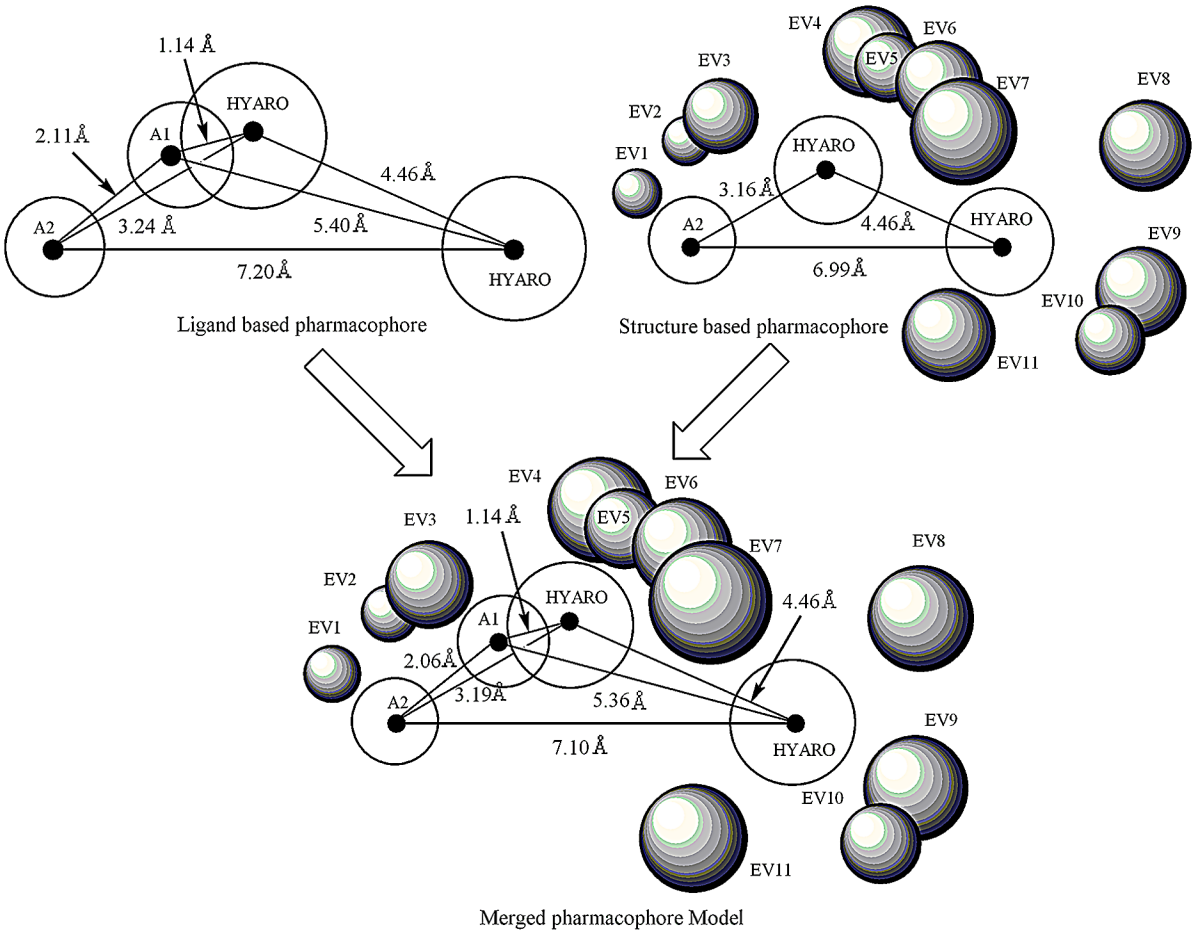


EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

training and test sets AIs. The LB model could predict 90% molecules, whereas the SB model and the merged model could predict 100% test set AIs. The merged pharmacophore model was able to predict these inhibitors and used to determine the active and the inactive compounds. This approach could be very useful for identifying lead discovery with improved inhibitory activity through 3D database search. Docking study analyzed the predictions for interactions between aromatase enzyme and two of the most potent inhibitors, i.e., verozole (**7**) and 3-[imidazol-1-yl-(4-nitrophenyl)-methyl]-chromen-4-one (**42**). The results showed that dock score and interaction energy values for the two nonsteroidal AIs were comparable to the preferred substrate androstenedione. Interaction of heteroatom with heme coordination was found to be implicated in each NSAIs. The docking study might be used to virtually evaluate the potency of new NSAIs. The pharmacophore requirements for potent NSAIs are shown in Figure 33.

Benzofuran derivatives were also found to be effective and potential aromatase inhibitors (Vinh *et al.* 1999; Saberi *et al*., 2006; Vinh *et al.*, 2000; Saberi, Shah, & Simons, 2005). The QSAR study (Nagar *et al*., 2009) revealed that the electron donating group at the C_1 atom (Figure 35) or the electron withdrawing groups at R′′′, hydrogen bond accepting character of the O_7 atom and the CN group substitution at the C_{14} atom as well as the decreasing steric effect might influence anti-aromatase activity. The pharmacophore

Figure 33. The ligand-based and structure-based pharmacophore requirements for potent NSAIs

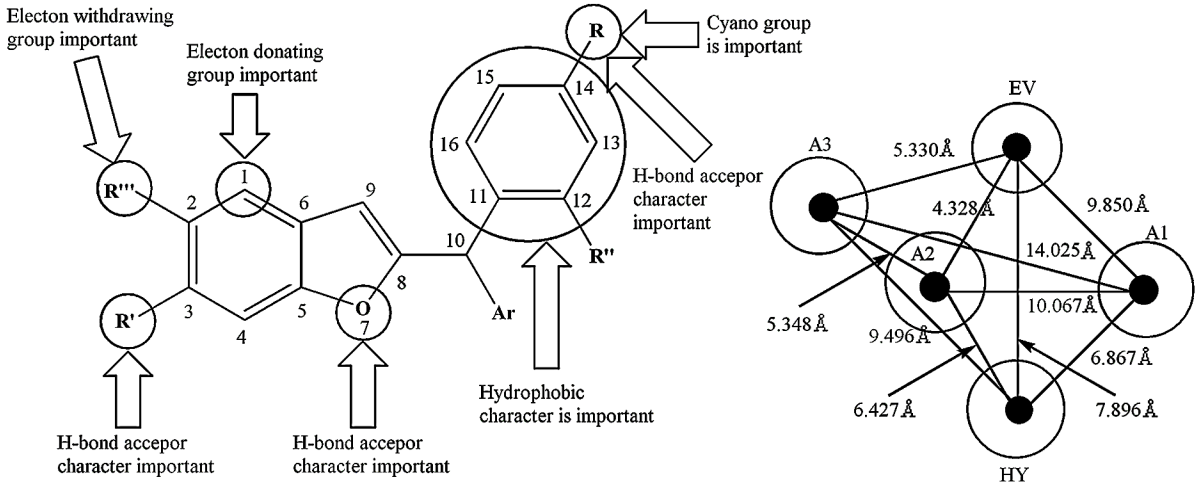


EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

mapping study of these molecules was well supported by the QSAR data. The best pharmacophore model showed the importance of three hydrogen bond acceptors, one hydrophobic feature and one excluded volume. The hydrogen bond acceptor features were separated from a distance of 5.348 Å, 10.367 Å and 14.025 Å respectively, whereas the hydrophobic feature was located 9.496 Å, 6.867 Å and 6.427 Å away from the hydrogen bond acceptor features, respectively. The distances between the excluded volume and the three hydrogen bond acceptor features and one hydrophobic feature were 5.330 Å, 4.328 Å, 9.850 Å and 7.896 Å, respectively. The pharmacophore mapping study suggested that hydrogen bond acceptor region at atoms C_3, O_7 and C_{14} was important for the activity, which was supposed to form the hydrogen bonding interaction with the receptor. A hydrophobic character was necessary for the aromatase inhibitory activity. The structural requirements of these benzofurans for promising anti-aromatase activity are shown in Figure 34.

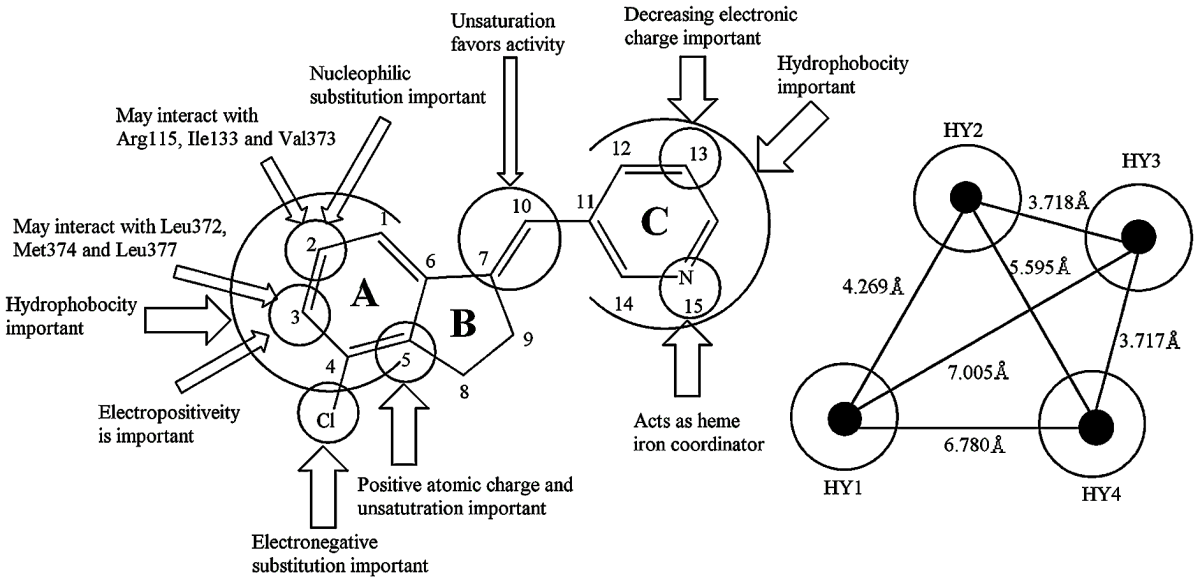
Benzcyclo derivatives were found to be effective aromatase inhibitors (Ulmschneider *et al*., 2005). The 2D-QSAR study (Nagar & Saha, 2010a) suggested that the nucleophilic substitution at the C_2 atom (Figure 35), the electronegative substitution at the C_4 atom, the positive atomic charge at the C_5 atom, the decreasing electronic charge at the atom number 13 and the increasing hydrophobicity of these molecules were important for anti-aromatase activity. The presence of the nitrogen atom at the C ring might have close interaction with Fe^{3+} of the heme to block the active site (Saberi, Shah, & Simons, 2005). Polar groups were found to be effective due to hydrophobic interactions in the active site cavity with the nonpolar amino acids like Ala306, Val370, Leu372, Val373, Met374, Leu377 and Leu477 (Karkola & Wahala, 2009; Roy & Roy, 2010b). The CoMFA study suggested that presence of the electropositive groups at C_3, C_5 and C_{13} positions favored aromatase inhibitory activity. Resonance effects due to the pattern of rings B and C as well as unsaturation at the C_5 atom may impart positive charge field at these positions. It was supported by the docking studies where it was observed that the electropositive substituent at the C_3 atom might have binding interactions with residues, like Leu372, Met374 and Leu377 of aromatase enzyme (Ghosh *et al*., 2009). Similarly, the presence of the electronegative group at the C_4 atom favored activity. The electronegative substituent at the C_4 atom was lying close to Leu477 residue of aromatase enzyme and had some negative electrostatic interaction with it. Again, the

Figure 34. Structural requirements of benzofurans for potent aromatase inhibitiory activity



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 35. Structural requirements of benzocyclo derivatives for potent anti-aromatase activity

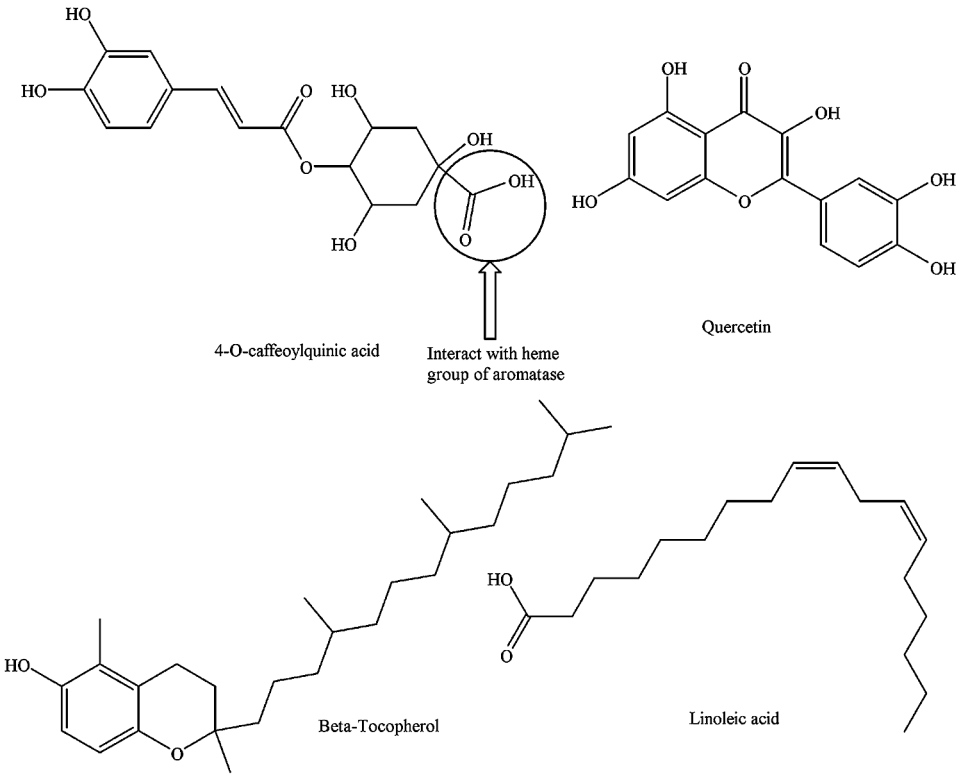


presence of the nitrogen atom at the heterocyclic C ring might have interaction with Fe^{3+} of the heme as well as the residue Ala306 in the active site of aromatase enzyme (Saberi, Shah, & Simons, 2005). The CoMSIA study supported similar pattern of the electrostatic interaction of the CoMFA model. In addition to electrostatic interactions at atoms C_3 and C_4, CoMSIA study suggested that electronegative substituent at the C_2 atom was important as that might have some binding interactions with residues, like Arg115, Ile133 and Val373 at the active site of aromatase enzyme. The electronegative substituent at the C_2 atom might bind with the receptor through H-bonding with Val373 and bump interaction with Arg115. Hydrophobicity was also playing a vital role to the aromatse inhibition. The polar groups at the C_3 atom might have hydrophobic interaction with the non-polar amino acid residues, like Met374 or Leu372 (Karkola & Wahala, 2009). The C_4 atom was also essential for the hydrophobic interactions with non-polar amino acid, Leu477. As the electronegative substitution at C_4 atom was also observed, hence, it might be inferred that the electronegative substituent at C_4 atom was more preferable rather than the hydrophobic interaction (Sun *et al*., 2010). The steric influence at atoms C_1, C_2, C_{12} and C_{13} was found to be favorable. The six member aromatic B ring is found to be more favorable than the five member ring system due to mesomeric effect. The unsaturation between atoms C_{10} and C_7 decreased steric hindrance and is favorable for the activity (Roy & Roy, 2010b). Pharmacophore mapping suggested that the four hydrophobic features (HY1, HY2, HY3 and HY4) were important for enzyme inhibition. HY1 feature was located 4.269 Å, 7.005 Å and 6.780 Å away from HY2, HY3 and HY4 features respectively, whereas the HY2 feature was located 3.718Å and 5.595Å away from HY3 and HY4 features, respectively. The distance between HY3 feature and HY4 feature was 3.717Å. The aromatic rings and aliphatic side chain in the ring A behaved as promising hydrophobic regions. The hydrophobic substitutions at the C_2 atom or the C_3 atom increased inhibitory activity by increasing the binding interaction with amino acid residues in the active site cavity. These observations are well supported by the other modeling studies (Neves *et al*., 2009; Saberi *et al*., 2006). The structural requirements of these benzocyclo derivatives for better anti-aromatase activity are shown in Figure 35.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Low molecular weight compounds from wild mushrooms were docked in aromatase enzyme and the potential SAR studies for the compounds with the best estimated inhibition constants were discussed for each compound family (Froufe, Abreu, & Ferreira, 2011). The compound 4-O-caffeoylquinic was found to be the top-ranked potential inhibitors for aromatase enzyme. Each family of low molecular weight compounds from wild mushrooms, like phenolic acid derivatives (benzoic and cinnamic acid derivatives and flavonoids), vitamins (tocopherols and ascorbic acid), carotenoids, sugars and fatty acids were docked against aromatase. It was found that benzoic acid derivatives were appeared to have no significant inhibitory activity against aromatase. Apart from 4-O-caffeylquinic and 5-O-caffeylquinic, cinnamic acid derivatives were not found to have significant activity. In case of flavonoids, quercetin was found to be the best compound to inhibit aromatase. The top-ranked compound 4-O-caffeoylquinic occupied the space of the aromatic rings of androstenedione and quinic acid might have polar contacts with heme group of aromatase through the carboxylic acid moiety. The compound may mimic androstenedione by forming hydrogen bond with Arg115 and peptide bond between Val373 and Met374. This might be possible due to one oxygen atom of 4-O-caffeoylquinic which was positioned on the same space occupied by the oxygen atom of one of carbonyl group of androstenedione. In case of vitamins and sugars, β-tocopherol and linoleic acid, respectively were found to be active against aromatase as far as the estimated K_i (μM) value was concerned. Important structural features required for higher anti-aromatase activity are shown in Figure 36.

Figure 36. Important structural features of natural compounds from mushroom for anti-aromatase activity

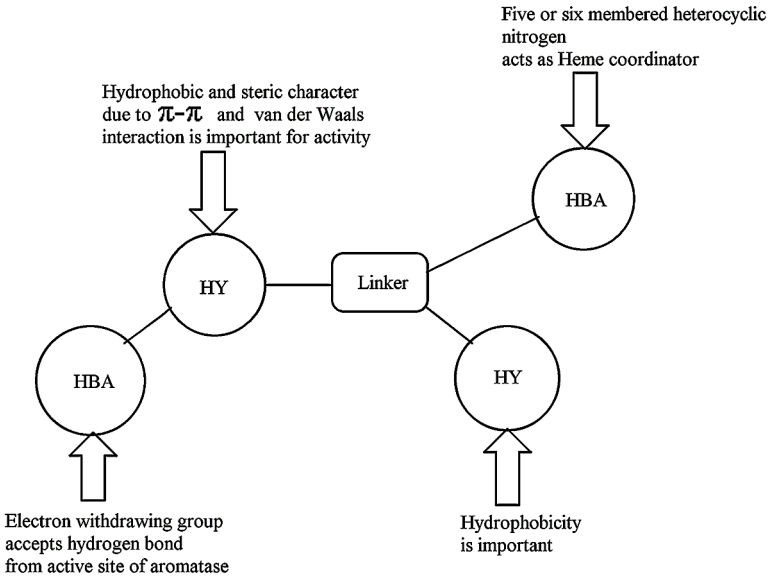


EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

FUTURE RESEARCH DIRECTIONS

Various molecular modeling studies, i.e. LBDD and SBDD approaches were taken into consideration in decades for developing better active and more specific aromatase inhibitors. As far as the structural aspects of the aromatase inhibitors are concerned, it is quite evident that NSAIs are less toxic than SAIs, since SAIs may cause androgenic and estrogenic adverse effects. Though it is very difficult to design and synthesize such molecules which have target specificity as well as lower toxicity, different molecular modeling studies may be capable of finding new small NSAIs within a short period of time. There are some chemometric methods which are still unexplored for modeling of aromatase inhibitors, like hologram QSAR (HQSAR), linear discriminant analysis (LDA)-based QSAR, group-based QSAR (GQSAR) etc. and SBDD approaches that include SB pharmacophore mapping, homology modeling, docking, molecular dynamics simulations, receptor-ligand site map analysis and *de novo* designing may open up a new vista for designing specific and novel hit aromatase inhibitors to fight against the dreaded disease of breast cancer. The HQSAR method is a fragment based LBDD approach that can be able to elucidate the important as well as unfavorable fragments of active and inactive molecules. The LDA-based QSAR on the available data of anti-aromatase molecules may be effective to know the important structural as well as physicochemical properties responsible for discriminating the active and the inactive anti-aromatase molecules. Similarly, GQSAR may able to identify new molecules of anti-aromatase activity. Apart from that, 3D-QSARs CoMFA as well as CoMSIA based studies followed by docking and virtual screening approach, lead identification and MD simulations may be possible routes for posulating new aromatase inhibitors. Besides, induced-fit docking method may also be effective for designing new aromatase inhibitors. The SBDD approach, *de novo* design may be a suitable choice for designing new potential anti-aromatase molecules.

Figure 37. Schematic representation for designing potential and selective non-steroidal aromatase inhibitors



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

CONCLUSION

In this chapter, different LBDD and SBDD approaches have presented to understand how these drug designing methods may be helpful for optimizing small molecule aromatase inhibitors. Though the modeling studies for designing anti-aromatase molecules have been continuing for more than two decades, it is a challenging task to design target specific and less toxic anti-aromatase cmpounds. Initially, it was a tedious job since less information was available about the enzyme structure and functions. Availability of the X-ray crystal structure as well as the discovery of different drug designing tools make the current process easier and more rational. In a nutshell, it is observed that for imparting anti-aromatase properties, molecules should have a five or six membered heterocyclic ring containing nitrogen atom so that it may coordinate with the heme portion of the aromatase enzyme. Apart from that, molecules should possess one or two negatively charged hydrogen bond acceptor features that can form hydrogen bonding with the amino acids present in the active site of the aromatase enzyme. Hydrophobicity and steric factors may also play crucial roles for higher activity as justified by the van der Waals interaction of molecules with the amino acid residues of the active site of the aromatase enzyme. Electrostatic and the π-π interactions with the catalytic amino acid are also found to be important for anti-aromatase properties. Some potential chemometric strategies may be undertaken to find hit molecules. Structure-based pharmacophore mapping followed by virtual screening and docking as well as MD simulations may be beneficial for exploring effective anti-aromatase drug. These observations may be well-supported and validated by experimental results. Ideally, the structural, physicochemical and electronic requirements of a non-steroidal molecule for imparting selective aromatase inhibition are schematically depicted in Figure 37. These findings may help to deign potential molecules that can reduce the rate of mortality in breast cancer in future.

REFERENCES

Accelrys Inc. (2011). Discovery studio 3.0. San Diego: Author.

Ahmed, K. M., Dong, S., Fan, M., & Li, J. J. (2006). Nuclear factor-kappaB p65 inhibits mitogen-activated protein kinase signalling pathway in radioresistant breast cancer cells. *Molecular Cancer Research*, *4*(12), 945–955. doi:10.1158/1541-7786.MCR-06-0291 PMID:17189385

Allen, N. E., Beral, V., Casabonne, D., Kan, S. W., Reeves, G. K., Brown, A., & Green, J. (2009). Moderate alcohol intake and cancer incidence in women. *Journal of the National Cancer Institute*, *101*(5), 296–305. doi:10.1093/jnci/djn514 PMID:19244173

Altundag, K., & Ibrahim, N. K. (2006). Aromatase inhibitors in breast cancer: An overview. *The Oncologist*, *11*(6), 553–562. doi:10.1634/theoncologist.11-6-553 PMID:16794235

Amarneh, B., Corbin, C. J., Peterson, J. A., Simpson, E. R., & Graham-Lorence, S. (1993). Functional domains of human aromatase cytochrome P450 characterized by linear alignment and site-directed mutagenesis. *Molecular Endocrinology (Baltimore, Md.)*, *7*(12), 1617–1624. PMID:8145767

Amato, E., Bankemper, T., Kidney, R., Do, T., Onate, A., & Shazna, T. et al. (2014). Investigation of fluorinated and bifunctionalized 3-phenylchroman-4-one (isoflavanone) aromatase inhibitors. *Bioorganic & Medicinal Chemistry*, *22*(1), 126–134. doi:10.1016/j.bmc.2013.11.045 PMID:24345481

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

American Cancer Society Cancer Facts & Figures 2013. (2013). Atlanta, GA: American Cancer Society.

Aparoy, P., Reddy, K. K., & Reddanna, P. (2012). Structure and ligand based drug design strategies in the development of novel 5-LOX inhibitors. *Current Medicinal Chemistry*, *19*(22), 3763–3778. doi:10.2174/092986712801661112 PMID:22680930

Ariazi, E. A., Ariazi, J. L., Cordera, F., & Jordan, V. C. (2006). Estrogen receptors as therapeutic targets in breast cancer. *Current Topics in Medicinal Chemistry*, *6*(3), 181–202. doi:10.2174/156802606776173483 PMID:16515478

Arumugam, A., Lissner, E. A., & Lakshmanaswamy, R. (2014). The role of hormones and aromatase inhibitors on breast tumor growth and general health in a postmenopausal mouse model. *Reproductive Biology and Endocrinology*, *12*(66), 1–13. PMID:25023195

Avis, N. E., Smith, K. W., Link, C. L., Hortobagyi, G. N., & Rivera, E. (2006). Factors associated with participation in breast cancer treatment clinical trials. *Journal of Clinical Oncology*, *24*(12), 1860–1867. doi:10.1200/JCO.2005.03.8976 PMID:16622260

Awasthi, M., Singh, S., Pandey, V. P., & Dwivedi, U. N. (2014). Molecular docking and 3D-QSAR-based virtual screening of flavonoids as potential aromatase inhibitors against estrogen-dependent breast cancer. *Journal of Biomolecular Structure & Dynamics*, 1–16. doi:10.1080/07391102.2014.912152 PMID:24702656

Bandi, P., Boone, M., Brinton, L., & Buchert, S. (2010). *Breast cancer facts & figures 2009-2010*. Atlanta, GA: American Cancer Society.

Banting, L., Nicholls, P. J., Shaw, M. A., & Smith, H. J. (1989). Recent developments in aromatase inhibition as a potential treatment for estrogen dependent breast cancer. *Progress in Medicinal Chemistry*, *26*, 253–298. doi:10.1016/S0079-6468(08)70242-X PMID:2690184

Banting, L., Smith, H. J., James, M., Jones, G., Nazareth, W., & Nicholls, P. J. et al. (1988). Structure-activity relationships for non-steroidal inhibitors of aromatase. *Journal of Enzyme Inhibition*, *2*(3), 215–229. doi:10.3109/14756368809040728 PMID:3241182

Barillari, C., Marcou, G., & Rognan, D. (2008). Hot-spots-guided receptor-based pharmacophores (HSPharm): A knowledge-based approach to identify ligand anchoring atoms in protein cavities and prioritize structure-based pharmacophores. *Journal of Chemical Information and Modeling*, *48*(7), 1396–1410. doi:10.1021/ci800064z PMID:18570371

Bayer, H., Batzl, C., Hartmann, R. W., & Mannschreck, A. (1991). New aromatase inhibitors. synthesis and biological activity of pyridyl-substituted tetralone derivatives. *Journal of Medicinal Chemistry*, *34*(9), 2685–2691. doi:10.1021/jm00113a004 PMID:1895288

Beatson, G. T. (1896). On the treatment of inoperable cases of carcinoma of the mamma: Suggestion for a new method of treatment, with illustrative cases 1. *Lancet*, *148*(3802), 101–107. doi:10.1016/S0140-6736(01)72307-0

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Bedard, P. L., Azambuja, E., & Cardoso, F. (2009). Beyond trastuzumab: Overcoming resistance to targeted HER2 therapy in breast cancer. *Current Cancer Drug Targets*, *9*(2), 148–162. doi:10.2174/156800909787581024 PMID:19275756

Beger, R. D., Buzatu, D. A., Wilkes, J. G., & Lay, J. O. Jr. (2001). ^{13}C NMR quantitative Spectrometric data-activity relationship (QSDAR) models of steroids binding the aromatase enzyme. *Journal of Chemical Information and Computer Sciences*, *41*(5), 1360–1366. doi:10.1021/ci010285e PMID:11604038

Benson, J. R., & Ravisekar, O. (2007). Aromatase inhibitors for treatment of breast cancer. *Current Cancer Therapy Reviewes*, *3*(1), 67–79. doi:10.2174/157339407780126656

Bhatnagar, A. S., Hausler, A., Schieweck, K., Lang, M., & Bowman, R. (1990). Highly selective inhibition of estrogen biosynthesis by CGS 20267, a new non-steroidal aromatase inhibitor. *The Journal of Steroid Biochemistry and Molecular Biology*, *37*(6), 1021–1027. doi:10.1016/0960-0760(90)90460-3 PMID:2149502

Bohm, H. J. (1992). The computer program LUDI: A new method for the de novo design of enzyme inhibitors. *Journal of Computer-Aided Molecular Design*, *6*(1), 61–78. doi:10.1007/BF00124387 PMID:1583540

Bonfield, K., Amato, E., Bankemper, T., Agard, H., Steller, J., Keeler, J. M., Roy, D., McCallum, A., Paula, S., & Ma, L. (2012). Development of a new class of aromatase inhibitors: Design, synthesis and inhibitory activity of 3-phenylchroman-4-one (isoflavanone) derivatives. *Bioorganic and Medicinal Chemistry letters, 20*, 2603-2613.

Brenk, R., & Klebe, G. (2006). 'Hot spot' analysis of protein-binding sites as a prerequisite for structure-based virtual screening and lead optimization. In T. Langer & R. D. Hoffmann (Eds.), *Pharmacophores and pharmacophore searches* (pp. 171–192). Weinheim, Germany: Wiley-VCH. doi:10.1002/3527609164.ch8

Brodie, A., Lu, Q., Liu, Y., & Long, B. (1999). Aromatase and its inhibitors. *The Journal of Steroid Biochemistry and Molecular Biology*, *69*(2), 205–210. doi:10.1016/S0960-0760(99)00051-5 PMID:10418994

Brodie, A., Sabnis, G., & Jelovac, D. (2006). Aromatase and breast cancer. *The Journal of Steroid Biochemistry and Molecular Biology*, *102*(1-5), 97–102. doi:10.1016/j.jsbmb.2006.09.002 PMID:17113978

Brodie, A. M., & Njar, V. C. (1996). Aromatase inhibitors and breast cancer. *Seminars in Oncology*, *23*(4-9), 10–20. PMID:8824460

Brodie, A. M., & Njar, V. C. (1998). Aromatase inhibitors in advanced breast cancer: Mechanism of action and clinical implications. *The Journal of Steroid Biochemistry and Molecular Biology*, *66*(1-2), 1–10. doi:10.1016/S0960-0760(98)00022-3 PMID:9712406

Brodie, A. M., & Njar, V. C. (2000). Aromatase inhibitors and their application in breast cancer treatment. *Steroids*, *65*(4), 171–179. doi:10.1016/S0039-128X(99)00104-X PMID:10713305

Brueggemeier, R. W. (1990). Biochemical and molecular aspects of aromatase. *Journal of Enzyme Inhibition*, *4*(2), 101–111. doi:10.3109/14756369009040731 PMID:2098516

Brueggemeier, R. W. (1994). Aromatase inhibitors mechanisms of steroidal inhibitors. *Breast Cancer Research and Treatment*, *30*(1), 31–42. doi:10.1007/BF00682739 PMID:7949203

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Brueggemeier, R. W., Floyd, E. E., & Counsell, R. E. (1978). Synthesis and biochemical evaluation of inhibitors of estrogen biosynthesis. *Journal of Medicinal Chemistry*, *21*(10), 1007–1011. doi:10.1021/jm00208a002 PMID:722711

Brueggemeier, R. W., Hackett, J. C., & Diaz-Cruz, E. S. (2005). Aromatase inhibitors in the treatment of breast cancer. *Endocrinology Review*, *26*(3), 331–345. doi:10.1210/er.2004-0015 PMID:15814851

Brufsky, A. M., & Glueck,. (2014). Current approaches and emerging directions in HER2-resistant breast cancer. *Basic and Clinical Research*, *8*, 109–118. doi:10.4137/BCBCR.S9453 PMID:25125981

Bubert, C., Woo, L. W. L., Sutcliffe, O. B., Mahon, M. F., Chander, S. K., & Purohit, A. et al. (2008). Synthesis, of aromatase inhibitors and dual aromatase steroid sulfatase inhibitors by linking an arylsulfamate motif to 4-(4H-1, 2, 4-triazol-4-ylamino)benzonitrile: SAR, crystal structure, in vitro and in vivo activities. *ChemMedChem*, *3*(11), 1708–1730. doi:10.1002/cmdc.200800164 PMID:18816537

Bulun, S. E., Price, T. M., Aitken, J., Mahandroo, M. S., & Simpson, E. R. (1993). A link between breast cancer and local estrogen biosynthesis suggested by quantification of adipose tissue aromatase cytochrome P450 transcripts using competitive polymerase chain reaction after reverse transcription. *The Journal of Clinical Endocrinology and Metabolism*, *77*(6), 1622–1628. PMID:8117355

Bulun, S. E., Takayama, K., Suzuki, T., Sasano, H., Yilmaz, B., & Sebastian, S. (2004). Organization of the human placental aromatase P450 (CYP19) gene. *Seminars in Reproductive Medicine*, *22*(1), 5–9. doi:10.1055/s-2004-823022 PMID:15083376

AJCC Cancer Staging Manual. (2002). New York, NY: Springer Verlag.

Carmichael, P. L. (1998). Mechanisms of action of antiestrogens: Relevance to clinical benefits and risks. *Cancer Investigation*, *16*(8), 604–611. doi:10.3109/07357909809032891 PMID:9844621

Castellano, S., Stefancich, G., Ragno, R., Schewe, K., Santoriello, M., & Caroli, A. et al. (2008). CYP19 (aromatase): Exploring the scaffold flexibility for novel selective inhibitors. *Bioorganic & Medicinal Chemistry*, *16*(18), 8349–8358. doi:10.1016/j.bmc.2008.08.046 PMID:18782670

Cavalli, A., Bisi, A., Bertucci, C., Rosini, C., Paluszcak, A., & Gobbi, S. et al. (2005). Enantioselective nonsteroidal aromatase inhibitors identified through a multidisciplinary medicinal chemistry approach. *Journal of Medicinal Chemistry*, *48*(23), 7282–7289. doi:10.1021/jm058042r PMID:16279787

Cavalli, A., Greco, G., Novellino, E., & Recanatini, M. (2000). Linking CoMFA and protein homology models of enzyme-inhibitor interactions: An application to non-steroidal aromatase inhibitors. *Bioorganic & Medicinal Chemistry*, *8*(12), 2771–2780. doi:10.1016/S0968-0896(00)00203-0 PMID:11131168

Cavalli, A., & Recanatini, M. (2002). Looking for selectivity among cytochrome P450s inhibitors. (2002). *Journal of Medicinal Chemistry*, *45*(2), 251–254. doi:10.1021/jm015567k PMID:11784128

Chemical Computing Group. (n.d.). Retrieved from www.chemcomp.com

Chen, J., & Lai, L. (2006). Pocket v.2: Further developments on receptor-based pharmacophore modeling. *Journal of Chemical Information and Modeling*, *46*(6), 2684–2691. doi:10.1021/ci600246s PMID:17125208

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chen, S. (1998). Aromatase and breast cancer. *Frontiers in Bioscience*, *3*, 922–933. PMID:9696881

Chen, S., Cho, M., Karlsberg, K., Zhou, D., & Yuan, Y. C. (2004). Biochemical and biological characterization of a novel anti-aromatase coumarin derivative. *The Journal of Biological Chemistry*, *279*(6), 48071–48078. doi:10.1074/jbc.M406847200 PMID:15358790

Cocconi, G. (1994). First generation aromatase inhibitors- aminoglutethimide and testololactone. *Breast Cancer Research and Treatment*, *30*(1), 57–80. doi:10.1007/BF00682741 PMID:7949205

Cole, P. A., & Robinson, C. H. (1990). Mechanism and inhibition of cytochrome P450 aromatase. *Journal of Medicinal Chemistry*, *33*(11), 2933–2942. doi:10.1021/jm00173a001 PMID:2231592

Colozza, M., Califano, R., Minenza, E., Dinh, P., & Azambuja, E. (2008). Aromatase inhibitors: A new reality for the adjuvant endocrine treatment of early-stage breast cancer in postmenopausal women. *Mini Reviews in Medicinal Chemistry*, *8*(6), 564–574. doi:10.2174/138955708784534472 PMID:18537711

Covey, D. F. (1988). Aromatase inhibitors: Specific inhibitors of oestrogen biosynthesis. In D. Berg & M. Plempel (Eds.), *Sterol biosynthesis inhibitors* (pp. 534–571). Chichester, UK: Ellis Horwood.

Dai, Y., Wang, Q., Zhang, X., Jia, S., Zheng, H., Feng, D., & Yu, P. (2010). Molecular docking and QSAR study on steroidal compounds as aromatase Inhibitors. *European Journal of Medicinal Chemistry*, *45*(12), 5612–5620. doi:10.1016/j.ejmech.2010.09.011 PMID:20926163

Davis, S. R., Guav, A. T., Shifren, J. L., & Mazer, N. A. (2004). Endocrine aspects of female sexual dysfunction. *Journal of Sexual Medicine*, *1*(1), 82–86. doi:10.1111/j.1743-6109.2004.10112.x PMID:16422987

Demers, L. M. (1994). Effects of fadrazole (CGS 16949A) and letrozole (CGS 20267) on the inhibition of aromatase activity in breast cancer patients. *Breast Cancer Research and Treatment*, *30*(1), 95–102. doi:10.1007/BF00682744 PMID:7949208

Dickson, R. B., & Lippman, M. E. (1995). Molecular basis of the breast cancer. In The molecular basis of cancer (pp. 358-386). Philedalphia: WB Saunders.

Dixon, S. L., Smondyrev, A. M., Knoll, E. H., Rao, S. N., Shaw, D. E., & Friesner, R. A. (2006). PHASE: A new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. *Journal of Computer-Aided Molecular Design*, *20*(10-11), 647–671. doi:10.1007/s10822-006-9087-6 PMID:17124629

Domingo-Domenech, J., Pippa, R., Tapia, M., Gascon, P., Bachs, O., & Bosch, M. (2008). Inactivation of NF-kappaB by proteasome inhibition contributes to increased apoptosis induced by histone deacetylase inhibitors in human breast cancer cells. *Breast Cancer Research and Treatment*, *112*(1), 53–62. doi:10.1007/s10549-007-9837-8 PMID:18064564

Dutta, U., & Pant, K. (2008). Aromatase inhibitors: Past, present and future in breast cancer therapy. *Medical Oncology (Northwood, London, England)*, *25*(2), 113–124. doi:10.1007/s12032-007-9019-x PMID:17973095

Early Breast Cancer Trialists' Collaborative Group. (2000). Favorable and unfavorable effects on long-term survival of radiotherapy for early breast cancer: An overview of the randomized trials. *Lancet*, *355*(9217), 1757–1770. doi:10.1016/S0140-6736(00)02263-7 PMID:10832826

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Early Breast Cancer Trialists' Collaborative Group. (2005). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15 year survival: An overview of the randomized trials. *Lancet*, *365*(9472), 1687–1717. doi:10.1016/S0140-6736(05)66544-0 PMID:15894097

Eisen, A., Trudeau, M., Shelley, W., Messersmith, H., & Pritchard, K. I. (2008). Aromatase inhibitors in adjuvant therapy for hormone receptor positive breast cancer: A systematic review. *Cancer Treatment Reviews*, *34*(2), 157–174. doi:10.1016/j.ctrv.2007.11.001 PMID:18164821

Esteban, J. M., Warsi, Z., Haniu, M., Hall, P. F., Shively, J. E., & Chen, S. (1992). Detection of intratumoral aromatase in breast carcinomas, an immunohistochemical study with clinico-pathologic correlation. *American Journal of Pathology*, *140*(2), 337–343. PMID:1739127

Favia, A. D., Cavalli, A., Masetti, M., Carotti, A., & Recanatini, M. (2006). Three dimensional model of the human aromatase enzyme and density functional parameterization of the iron-containing protopophyrin IX for molecular dynamics study of heme-cysteinato cytochromes. *Proteins*, *62*(4), 1074–1087. doi:10.1002/prot.20829 PMID:16395678

Fisher, B., Anderson, S., Bryant, J., Margolese, R. G., Deutsch, M., & Fisher, E. R. et al. (2002). Twenty-year follow-up of a randomized trial comparing total mastectomy, lumpectomy and lumpectomy plus irradiation for the treatment of invasive breast cancer. *The New England Journal of Medicine*, *347*(16), 1233–1241. doi:10.1056/NEJMoa022152 PMID:12393820

Froufe, H. J. C., Abreu, R. M. V., & Ferreira, I. C. F. R. (2011). Using molecular docking to investigate the anti-breast cancer activity of low molecular weight compounds present on wild mushrooms. *SAR and QSAR in Environmental Research*, *22*(3-4), 315–328. doi:10.1080/1062936X.2011.569897 PMID:21598196

Furet, P., Batzl, C., Bhatnagar, A., Francotte, E., Rihs, G., & Lang, M. (1993). Aromatase inhibitors: Synthesis, biological activity, and binding mode of azole-type compounds. *Journal of Medicinal Chemistry*, *36*(10), 1393–1400. doi:10.1021/jm00062a012 PMID:8496907

Gartner, C. A., Thompson, S. J., Rettie, A. E., & Nelson, S. D. (2001). Human aromatase in high yield and purity by perfusion chromatography and its characterization by different spectroscopy and mass spectrometry. *Protein Expression and Purification*, *22*(3), 443–454. doi:10.1006/prep.2001.1464 PMID:11483007

Geisler, J. (2003). Aromatase inhibitors and inactivators for the treatment of postmenopausal breast cancer: A review. *Current Medicinal Chemistry - Immunology, Endocrine & Metabolic Agents*, *3*(3), 261–276. doi:10.2174/1568013033483339

Geisler, J., & Lonning, P. E. (2005). Aromatase inhibition: Translation into a successful therapeutic approach. *Clinical Cancer Research*, *11*(8), 2809–2821. doi:10.1158/1078-0432.CCR-04-2187 PMID:15837728

Ghosh, D., Griswold, J., Erman, M., & Pangborn, W. (2009). Structural basis for androgen specificity and oestrogen synthesis in human aromatase. *Nature*, *457*(7226), 219–223. doi:10.1038/nature07614 PMID:19129847

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Ghosh, D., Griswold, J., Erman, M., & Pangborn, W. (2010). X-ray structure of human aromatase reveals an androgen-specific active site. *The Journal of Steroid Biochemistry and Molecular Biology*, *118*(4-5), 197–202. doi:10.1016/j.jsbmb.2009.09.012 PMID:19808095

Ghosh, D., Jiang, W., Lo, J., & Egbuta, C. (2011). Higher order organization of human placental aromatase. *Steroids*, *76*(8), 753–758. doi:10.1016/j.steroids.2011.02.030 PMID:21392520

Gligorov, J., & Lotz, J. P. (2008). Optimal treatment strategies in postmenopausal women with hormone receptor-positive and HER2-negative metastatic breast cancer. *Breast Cancer Research and Treatment*, *112*(1), 53–66. doi:10.1007/s10549-008-0232-x PMID:19101794

Gobbi, S., Cavalli, A., Bisi, A., & Recanatini, M. (2008). From nonsteroidal aromatase inhibitors to multifunctional drug candidates: Classic and innovative strategies for the treatment of breast cancer. *Current Topics in Medicinal Chemistry*, *8*(10), 869–887. doi:10.2174/156802608784911590 PMID:18673172

Gobbi, S., Cavalli, A., Negri, M., Schewe, K. E., Belluti, F., & Piazzi, L. et al. (2007). Imidazolylmethylbenzophenones as highly potent aromatase inhibitors. *Journal of Medicinal Chemistry*, *50*(15), 3420–3422. doi:10.1021/jm0702938 PMID:17585752

Gobbi, S., Cavalli, A., Rampa, A., Belluti, F., Piazzi, L., & Paluszcak, A. et al. (2006). Lead optimization providing a series of flavone derivatives as potent nonsteroidal inhibitors of the cytochrome P450 aromatase enzyme. *Journal of Medicinal Chemistry*, *49*(15), 4777–4780. doi:10.1021/jm060186y PMID:16854084

Gobbi, S., Zimmer, C., Belluti, F., Rampa, A., Hartmann, R. W., Recanatini, M., & Bisi, A. (2010). Novel highly potent and selective nonsteroidal aromatase inhibitors: Synthesis, biological evaluation and structure-activity relationships investigation. *Journal of Medicinal Chemistry*, *53*(14), 5347–5351. doi:10.1021/jm100319h PMID:20568782

Goldblatt, E. M., Erickson, P. A., Gentry, E. R., Gryaznov, S. M., & Herbert, B. S. (2009). Lipid-conjugated telomerase template antagonists sensitize resistant HER2-positive breast cancer cells to trastuzumab. *Breast Cancer Research and Treatment*, *118*(1), 21–32. doi:10.1007/s10549-008-0201-4 PMID:18853252

Gupta, R., Jindal, D. P., Jit, B., Narang, G., Palusczak, A., & Hartmann, R. W. (2004). Synthesis and evaluation of a dimer of 2-(4-pyridylmethyl)-1-indanone as a novel nonsteroidal aromatase inhibitor. *Archiv der Pharmazie*, *337*(7), 398–401. doi:10.1002/ardp.200400853 PMID:15237390

Hackett, J. C., Kim, Y. W., Su, B., & Brueggemeier, R. W. (2005). Synthesis and characterization of azole isoflavone inhibitors of aromatase. *Bioorganic & Medicinal Chemistry*, *13*(12), 4063–4070. doi:10.1016/j.bmc.2005.03.050 PMID:15911319

Harada, N. (1997). Aberrant expression of aromatase in breast cancer tissues. *The Journal of Steroid Biochemistry and Molecular Biology*, *61*(3-6), 175–184. doi:10.1016/S0960-0760(97)80010-6 PMID:9365188

Hartmann, R. W., Bayer, H., & Grun, G. (1994). Aromatase inhibitors. syntheses and structure-activity studies of novel pyridyl-substituted indanones, indans, and tetralins. *Journal of Medicinal Chemistry*, *37*(9), 1275–1281. doi:10.1021/jm00035a007 PMID:8176705

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Hartmann, R. W., Bayer, H., Grun, G., Sergejew, T., Bartz, U., & Mitrenga, M. (1995). Pyridyl-substituted tetrahydrocyclopropa[alnaphthalenes: Highly active and selective inhibitors of P450 arom. *Journal of Medicinal Chemistry*, *38*(12), 2103–2111. doi:10.1021/jm00012a009 PMID:7783141

Hartmann, R. W., Frotscher, M., Ledergerber, D., Wächter, G. A., Grün, G. L., & Sergejew, T. F. (1996). synthesis and evaluation of azole-substituted tetrahydronaphthalenes as inhibitors of P450 arom, P450 17, and P450 TxA_2. *Archiv der Pharmazie*, *329*(5), 251–261. doi:10.1002/ardp.19963290506 PMID:8779634

Hartmann, R. W., Paiusczak, A., Lacan, F., & Ruzziconi, R. (2004). CYP 17 and CYP 19 inhibitors. evaluation of fluorine effects on the inhibiting activity of regioselectively fluorinated 1-(Naphthalen-2-ylmethyl) imidazoles. *Journal of Enzyme Inhibition and Medicinal Chemistry*, *19*(2), 145–155. doi:10.1080/1475636042000196222 PMID:15449729

Hoffken, K. (1993). Experience with aromatase inhibitors in the treatment of advanced breast cancer. *Cancer Treatment Reviews*, *19*(Suppl. B), 37–44. doi:10.1016/0305-7372(93)90006-D PMID:8481933

Hollander, P. D., Savage, M. I., & Brown, P. H. (2013). Targeted therapy for breast cancer prevention. *Frontiers in Oncology*, *3*(250), 1–15. PMID:23373009

Hong, Y., Yu, B., Sherman, M., Yuan, Y. C., Zhou, D., & Chen, S. (2007). Molecular basis for the aromatization reaction and exemestane-mediated irreversible inhibition of human aromatase. *Molecular Endocrinology (Baltimore, Md.)*, *21*(2), 401–414. doi:10.1210/me.2006-0281 PMID:17095574

Howe, L. R., Subbaramaiah, K., Patel, J., Masferrer, J. L., Deora, A., & Hudis, C. et al. (2002). Celecoxib, a selective cyclooxygenase 2 inhibitor, protects against human epidermal growth factor receptor2 (HER-2)/neu-induced breast cancer. *Cancer Research*, *62*(19), 5405–5407. PMID:12359744

Howell, A. (2005). New developments in the treatment of postmenopausal breast cancer. *Trends in Endocrinology and Metabolism*, *16*(9), 420–428. doi:10.1016/j.tem.2005.09.003 PMID:16213745

Howlader, N., Altekruse, S. F., Li, C. I., Chen, V. W., Clarke, C. A., Ries, L. A., & Cronin, K. A. (2014). US incidence of breast cancer subtypes defined by joint hormone receptor and HER2 status. *Journal of the National Cancer Institute*, *106*(5), 1–8. doi:10.1093/jnci/dju055 PMID:24777111

Hurtado, A., Holmes, K. A., Geistlinger, T. R., Hutcheson, I. R., Nicholson, R. I., & Brown, M. et al. (2008). Regulation of ErbB2 by oestrogen receptor-PAX2 determine response to tamoxifen. *Nature*, *456*(7222), 663–666. doi:10.1038/nature07483 PMID:19005469

Jackson, T., Woo, L. W., Trusselle, M. N., Purohit, A., Reed, M. J., & Potter, B. V. (2008). Non-steroidal aromatase inhibitors based on a biphenyl scaffold: Synthesis, in vitro SAR, and molecular modelling. *ChemMedChem*, *3*(4), 603–618. doi:10.1002/cmdc.200700266 PMID:18236493

Jacobs, C., Frotscher, M., Dannhardt, G., & Hartmann, R. W. (2000). 1-Imidazolyl(alkyl)-substituted di- and tetrahydroquinolines and analogs: Syntheses and evaluation of dual inhibitors of thromboxane A2 synthase and aromatase. *Journal of Medicinal Chemistry*, *43*(9), 1841–1851. doi:10.1021/jm991180u PMID:10794700

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

James, V. H., Mcneill, J. M., Lai, L. C., Newton, C. J., Ghilchik, M. W., & Reed, M. J. (1987). Aromatase activity in normal breast and breast tumor tissues: In vivo and in vitro studies. *Steroids*, *50*(1-3), 269–279. doi:10.1016/0039-128X(83)90077-6 PMID:3509763

Jiang, W., & Ghosh, D. (2012). Motion and flexibility in human cytochrome P450 aromatase. *PLoS ONE*, *7*(2), e32565. doi:10.1371/journal.pone.0032565 PMID:22384274

Johnston, J. O., & Metcalf, B. W. (1984). Aromatase: A target enzyme in breast cancer. In P. Sunkara (Ed.), *Novel approaches to cancer chemotherapy* (pp. 307–328). New York: Academic Press. doi:10.1016/B978-0-12-676980-7.50014-7

Jones, C. D., Winter, M. A., Hirsch, K. S., Stamm, N., Taylor, H. M., & Holden, H. E. et al. (1990). Estrogen synthetase inhibitors 2. Comparison of the in vitro aromatase inhibitory activity for a variety of nitrogen heterocycles substituted with diarylmethane and diarylmethanol groups. *Journal of Medicinal Chemistry*, *33*(1), 416–429. doi:10.1021/jm00163a065 PMID:2296032

Jones, G., & Willet, P. (2000). GASP: Genetic algorithm superimposition program. In O. F. Guner (Ed.), *Pharmacophore perception, development, and use in drug design* (pp. 85–106). La Jolla, CA: International University Line.

Jordan, V. C. (1995). Tamoxifen: Toxicities and drug resistance during the treatment and prevention of breast cancer. *Annual Review of Pharmacology and Toxicology*, *35*(1), 195–211. doi:10.1146/annurev.pa.35.040195.001211 PMID:7598491

Kapetanovic, I. M. (2008). Computer-aided drug discovery and development (CADDD): In silico-chemico-biological approach. *Chemico-Biological Interactions*, *171*(2), 165–176. doi:10.1016/j.cbi.2006.12.006 PMID:17229415

Karjalainen, A., Kalapudas, A., Sodervall, M., Pelkonen, O., & Lammintausta, R. (2000). Synthesis of new potent and selective aromatase inhibitors based on long chained diarylalkylimidazole and diarylalkyltriazole molecule skeletons. *European Journal of Pharmaceutical Sciences*, *11*(2), 109–131. doi:10.1016/S0928-0987(00)00074-9 PMID:10915960

Karkola, S., & Wähälä, K. (2009). The binding of lignans, flavonoids and coumestrol to CYP450 aromatase: A molecular modeling study. *Molecular and Cellular Endocrinology*, *301*(1-2), 235–244. doi:10.1016/j.mce.2008.10.003 PMID:19000737

Kellis, J. T. Jr, & Vickery, L. E. (1987). Purification and characterization of human placental aromatase cytochrome P450. *The Journal of Biological Chemistry*, *262*(9), 4413–4420. PMID:3104339

Khodarahmi, G. A., Laughton, C. A., Smith, H. J., & Nicholls, P. J. (2001). Enantioselectivity of some 1-[(Benzofuran-2-yl) phenylmethyl] imidazoles as aromatase (P450arom) inhibitors. *Journal of Enzyme Inhibition*, *16*(5), 401–416. doi:10.1080/14756360109162389 PMID:11916146

Kim, Y. W., Hackett, J. C., & Brueggemeier, R. W. (2004). Synthesis and aromatase inhibitory activity of novel pyridine-containing isoflavones. *Journal of Medicinal Chemistry*, *47*(16), 4032–4040. doi:10.1021/jm0306024 PMID:15267241

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Koymans, L. M. H., Moereels, H., & Bossche, H. V. (1995). A molecular model for the interaction between vorozole and other non-steroidal inhibitors and human cytochrome P450 19 (P450 aromatase). *The Journal of Steroid Biochemistry and Molecular Biology*, *53*(1-6), 191–197. doi:10.1016/0960-0760(95)00033-V PMID:7626453

Kroemer, R. T. (2007). Structure-based drug design: Docking and scoring. *Current Protein & Peptide Science*, *8*(4), 312–328. doi:10.2174/138920307781369382 PMID:17696866

Labrie, F., Belanger, A., Cusan, L., & Candes, B. (1997). Physiological changes in dehydroepiandrosterone are not reflected by serum levels of active androgens and estrogens but of their metabolites: Intracrinology. *The Journal of Clinical Endocrinology and Metabolism*, *82*(8), 2403–2409. doi:10.1210/jcem.82.8.4161 PMID:9253308

Lang, M., Batzl, C., Furet, P., Bowman, R., Hausler, A., & Bhatnagar, A. (1993). Structure-activity relationships and binding model of novel aromatase inhibitors. *The Journal of Steroid Biochemistry and Molecular Biology*, *44*(4-6), 421–428. doi:10.1016/0960-0760(93)90245-R PMID:8476755

Le Bail, J. C., Pouget, C., Fagnere, C., Basly, J. P., Chulia, A. J., & Habrioux, G. (2001). Chalcones are potent inhibitors of aromatase and 17β-hydroxysteroid dehydrogenase activities. *Life Sciences*, *68*(7), 751–761. doi:10.1016/S0024-3205(00)00974-7 PMID:11205867

Le Borgne, M., Marchand, P., Nourrisson, M. R., Loquet, D., Palzer, M., Le Baut, G., & Hartmann, R. W. (2007). Synthesis and biological evaluation of 3-(azolylmethyl)-1H-indoles and 3-(alpha-azolylbenzyl)-1H-indoles as selective aromatase inhibitors. *Journal of Enzyme Inhibition and Medicinal Chemistry*, *22*(5), 667–676. doi:10.1080/14756360701652658 PMID:18035835

Le Brogne, M., Marchand, P., Delevove-Seiller, B., Robert, J. M., Le Baut, G., Hartmann, R. W., & Palzer, M. (1999). New selective nonsteroidal aromatase inhibitors: Synthesis and inhibitory activity of 2,3 or 5-(alpha-azolylbenzyl)-1H-indoles. *Bioorganic & Medicinal Chemistry Letters*, *9*(3), 333–336. doi:10.1016/S0960-894X(98)00737-9 PMID:10091679

Le'ze´, M. P., Le Borgne, M., Marchand, P., Loquet, D., Kogler, M., & Le Bauta, G. et al. (2004). 2- and 3-[(aryl)(azolyl)methyl]indoles as potential non-steroidal aromatase inhibitors. *Journal of Enzyme Inhibition and Medicinal Chemistry*, *19*(6), 549–557. doi:10.1080/14756360400004631 PMID:15662958

Le'ze´, M. P., Le Borgne, M., Pinson, P., Palusczak, A., Duflos, M., Le Bauta, G., & Hartmann, R. W. (2006). Synthesis and biological evaluation of 5-[(aryl)(1H-imidazol-1-yl)methyl]-1H-indoles: Potent and selective aromatase inhibitors. *Bioorganic & Medicinal Chemistry Letters*, *16*(5), 1134–1137. doi:10.1016/j.bmcl.2005.11.099 PMID:16380254

Le'ze´, M. P., Palusczak, A., Hartmann, R. W., & Le Borgne, M. (2008). Synthesis of 6- or 4-functionalized indoles via a reductive cyclization approach and evaluation as aromatase inhibitors. *Bioorganic & Medicinal Chemistry Letters*, *18*(16), 4713–4715. doi:10.1016/j.bmcl.2008.06.094 PMID:18640836

Lengauer, T., & Rarey, M. (1996). Computational methods for biomolecular docking. *Current Opinion in Structural Biology*, *6*(3), 402–406. doi:10.1016/S0959-440X(96)80061-3 PMID:8804827

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Leonetti, F., Favia, A., Rao, A., Aliano, R., Paluszcak, A., Hartmann, R. W., & Carotti, A. (2004). Design, synthesis, and 3D-QSAR of novel potent and selective aromatase inhibitors. *Journal of Medicinal Chemistry*, *47*(27), 6792–6803. doi:10.1021/jm049535j PMID:15615528

Lipton, A., Harvey, H. A., & Santen, R. J. (1982). Closing remarks to the conference, aromatase: New perspectives for breast cancer. *Cancer Research*, *42*(8), 3468s. PMID:7083210

Liu, F., Xie, Z. H., Cai, G. P., & Jiang, Y. Y. (2007). The effect of surviving on multidrug resistance mediated by P-glycoprotein in MCF-7 and its adriamycin resistant cells. *Biological & Pharmaceutical Bulletin*, *30*(12), 2279–2283. doi:10.1248/bpb.30.2279 PMID:18057712

Lyman, G. H., Giuliano, A. E., & Somerfield, M. R. (2005). American society of clinical oncology guideline recommendations for seminal lymph node biopsy in early-stage breast cancer. *Journal of Clinical Oncology*, *23*(30), 7703–7720. doi:10.1200/JCO.2005.08.001 PMID:16157938

Marchand, P., Le Borgne, M., Palzer, M., Le Bauta, G., & Hartmann, R. W. (2003). Preparation and pharmacological profile of 7-(alpha-azolylbenzyl)-1H-indoles and indolines as new aromatase inhibitors. *Bioorganic & Medicinal Chemistry Letters*, *13*(9), 1553–1555. doi:10.1016/S0960-894X(03)00182-3 PMID:12699753

Martin, Y. C. (2000). DISCO: what we did right and what we missed. In O. F. Guner (Ed.), *Pharmacophore perception, development, and use in drug design* (pp. 51–66). La Jolla, CA: International University Line.

Mauri, D., Pavlidis, N., & Loannidis, J. P. (2005). Neoadjuvant versus adjuvant systemic treatment in breast cancer: A meta-analysis. *Journal of the National Cancer Institute*, *97*(3), 188–194. doi:10.1093/jnci/dji021 PMID:15687361

Mauri, D., Polyzos, N. P., Salanti, G., Pavlidis, N., & Ioannidis, J. P. (2008). Multiple treatments meta-analysis of chemotherapy and targeted therapies in advanced breast cancer. *Journal of the National Cancer Institute*, *100*(24), 1780–1791. doi:10.1093/jnci/djn414 PMID:19066278

May, F. E. B. (2014). Novel drugs that target the estrogen-related receptor alpha: Their therapeutic potential in breast cancer. *Cancer Management and Research*, *6*, 225–252. doi:10.2147/CMAR.S35024 PMID:24904222

McCullough, L. E., Eng, S. M., Bradshaw, P. T., Cleveland, R. J., Teitelbaum, S. L., Neugut, A. I., & Gammon, M. D. (2012). Fat or fit: The joint effects of physical activity, weight gain, and body size on breast cancer risk. *Cancer*, *118*(19), 4860–4868. doi:10.1002/cncr.27433 PMID:22733561

Michaelson, J., Satija, S., Moore, R., Weber, G., Halpern, E., & Garland, A. et al. (2002). The pattern of breast cancer screening utilization and its consequences. *Cancer*, *94*(1), 37–43. doi:10.1002/cncr.10154 PMID:11815958

Miller, K. D.(2012). Inhibiting telomerase to reverse trastuzumab(T) resistance in HER2+ breast cancer. *Cancer Research*, *72*(24 suppl).

Miller, W. R. (1997). Aromatase inhibitors and breast cancer. *Cancer Treatment Reviews*, *23*(3), 171–187. doi:10.1016/S0305-7372(97)90037-2 PMID:9251721

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Miller, W. R. (2003). Aromatase inhibitors: Mechanism of action and role in the treatment of breast cancer. *Seminars in Oncology, 30*(4-14), 3-11.

Miller, W. R., Bartlett, J., Brodie, A. M., Brueggemeier, R. W., di Salle, E., & Lonning, P. E. et al. (2008). Aromatase inhibitors: Are there differences between steroidal and nonsteroidal aromatase inhibitors and do they matter? *The Oncologist*, *13*(8), 829–837. doi:10.1634/theoncologist.2008-0055 PMID:18695261

Miller, W. R., & O'neill, J. (1987). The importance of local synthesis of estrogen within the breast. *Steroids*, *50*(4-6), 537–548. doi:10.1016/0039-128X(87)90037-7 PMID:3504615

Modi, S., Saura, C., Henderson, C., Lin, N. U., Mahtani, R., & Goddard, J. et al. (2013). A multicenter trial evaluating retaspimycin HCL (IPI-504) plus trastuzumab in patients with advanced or metastatic HER2-positive breast cancer. *Breast Cancer Research and Treatment*, *139*(1), 107–113. doi:10.1007/s10549-013-2510-5 PMID:23580070

Modi, S., Stopeck, A., Linden, H., Solit, D., Chandarlapaty, S., & Rosen, N. et al. (2011). HSP90 inhibition is effective in breast cancer: A phase II trial of tanespimycin (17-AAG) plus trastuzumab in patients with HER2-positive metastatic breast cancer progressing on trastuzumab. *Clinical Cancer Research*, *17*(15), 5132–5139. doi:10.1158/1078-0432.CCR-11-0072 PMID:21558407

Mohammed, H. A., Ba, L. A., Burkholz, T., Schumann, E., Diesel, B., & Zapp, J. et al. (2011). Facile synthesis of chrysin-derivatives with promising activities as aromatase inhibitors. *Natural Product Communications*, *6*(1), 31–34. PMID:21366040

Mohan, V., Gibbs, A. C., Cummings, M. D., Jaeger, E. P., & DesJarlais, R. L. (2005). Docking: Successes and challenges. *Current Pharmaceutical Design*, *11*(3), 323–333. doi:10.2174/1381612053382106 PMID:15723628

Morales, A., Buvat, J., Gooren, L. J., Guav, A. T., Kaufman, J. M., Tan, H. M., & Torres, L. O. (2004). Endocrine aspects of sexual dysfunction in men. *Journal of Sexual Medicine*, *1*(1), 69–81. doi:10.1111/j.1743-6109.2004.10111.x PMID:16422986

Muftuoglu, Y., & Mustata, G. (2010). Pharmacophore modeling strategies for the development of novel nonsteroidal inhibitors of human aromatase (CYP19). *Bioorganic & Medicinal Chemistry Letters*, *20*(10), 3050–3064. doi:10.1016/j.bmcl.2010.03.113 PMID:20413308

Murthy, J. N., Nagaraju, M., Sastry, G. M., Rao, A. R., & Sastry, G. N. (2006). Active site acidic residues and structural analysis of modeled human aromatase: A potential drug target for breast cancer. *Journal of Computer-Aided Molecular Design*, *19*(12), 857–870. doi:10.1007/s10822-005-9024-0 PMID:16741831

Murthy, J. N., Rao, A. R., & Sastry, G. N. (2004). Aromatase inhibitors: A new paradigm in breast cancer treatment. *Current Medicinal Chemistry. Anti-Cancer Agents*, *4*(6), 523–534. doi:10.2174/1568011043352669 PMID:15579017

Muti, P., Rogan, E., & Cavalieri, E. (2006). Androgens and estrogens in the etiology and prevention of breast cancer. *Nutrition and Cancer*, *56*(2), 247–252. doi:10.1207/s15327914nc5602_17 PMID:17474872

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Nagar, S., Islam, M. A., Das, S., Mukherjee, A., & Saha, A. (2008). Pharmacophore mapping of flavone derivatives for aromatase inhibition. *Molecular Diversity*, *12*(1), 65–76. doi:10.1007/s11030-008-9077-9 PMID:18506592

Nagar, S., Islam, M. A., Das, S., Mukherjee, A., & Saha, A. (2009). Pharmacophore searching of benzofuran derivatives for selective CYP19 aromatase inhibition. *Letters in Drug Design & Discovery*, *6*(1), 38–45. doi:10.2174/157018009787158607

Nagar, S., & Saha, A. (2010a). Exploring benzcyclo derivatives as potent aromatase inhibitors using ligand-based modeling studies. *European Journal of Medicinal Chemistry*, *45*(9), 4307–4315. doi:10.1016/j.ejmech.2010.06.033 PMID:20638757

Nagar, S., & Saha, A. (2010b). Modeling of diarylalkyl-imidazole and diarylalkyl-triazole derivatives as potent aromatase inhibitors for treatment of hormone-dependent cancer. *Journal of Computational Chemistry*, *31*(12), 2342–2353. PMID:20340106

Nagar, S., & Saha, A. (2010c). QSAR modeling and pharmacophore mapping of isoflavone derivatives for aromatase inhibitory activity. *International Journal of Pharmacy and Pharmaceutical Sciences*, *2*(3), 126–131.

Nagy, P. I., Tokarski, J., & Hopfinger, A. J. (1994). Molecular shape and QSAR analyses of a family of substituted dichlorodiphenyl aromatase inhibitors. *Journal of Chemical Information and Computer Sciences*, *34*(5), 1190–1197. doi:10.1021/ci00021a029 PMID:7962215

Nahta, R., Yuan, L. X., Zhang, B., Kobayashi, R., & Esteva, F. J. (2005). Insulin-like growth factor- I receptor/human epidermal growth factor receptor 2 heterodimerization contributes to trastuzumab resistance of breast cancer cells. *Cancer Research*, *65*(23), 11118–11128. doi:10.1158/0008-5472.CAN-04-3841 PMID:16322262

Nakshatri, H., Bhat-Nakshatri, P., Martin, D. A., Goulet, R. J. Jr, & Sledge, G. W. Jr. (1997). Constitutive activation of NF-kappaB during progression of breast cancer to hormone-dependent growth. *Molecular and Cellular Biology*, *17*(7), 3629–3639. PMID:9199297

Nantasenamat, C., Li, H., Mandi, P., Worachartcheewan, A., Monnor, T., Ayudhya, C. I. N., & Prachayasittikul, V. (2013). Exploring the chemical space of aromatase inhibitors. *Molecular Diversity*, *17*(4), 661–677. doi:10.1007/s11030-013-9462-x PMID:23857318

Nantasenamat, C., Worachartcheewan, A., Mandi, P., Monnor, T., Ayudhya, C. I. N., & Prachayasittikul, V. (2014). QSAR modeling of aromatase inhibition by flavonoids using machine learning approaches. *Chemical Papers*, *68*(5), 697–713. doi:10.2478/s11696-013-0498-2

Nantasenamat, C., Worachartcheewan, A., Prachayasittikul, S., Ayudhya, C. I. N., & Prachayasittikul, V. (2013). QSAR modeling of aromatase inhibitory activity of 1-substituted 1, 2, 3-triazole analogs of letrozole. *European Journal of Medicinal Chemistry*, *69*(1), 99–114. doi:10.1016/j.ejmech.2013.08.015 PMID:24012714

Narayana, B. L., Kishore, D. P., Balakumar, C., Rao, K. V., Kaur, R., & Rao, A. R. et al. (2012). Molecular modeling evaluation of non-steroidal aromatase inhibitors. *Chemical Biology & Drug Design*, *79*(5), 674–682. doi:10.1111/j.1747-0285.2011.01277.x PMID:22129073

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Neves, M. A., Dinis, T. C., Colombo, G., & Sae Melo, M. L. (2009a). An efficient steroid pharmacophore-based strategy to identify new aromatase inhibitors. *European Journal of Medicinal Chemistry*, *44*(10), 4121–4127. doi:10.1016/j.ejmech.2009.05.003 PMID:19500885

Neves, M. A., Dinis, T. C., Colombo, G., & Sae Melo, M. L. (2009b). Fast three dimensional pharmacophore virtual screening of new potent non-steroid aromatase inhibitors. *Journal of Medicinal Chemistry*, *52*(1), 143–150. doi:10.1021/jm800945c PMID:19072235

Numazawa, M., Shelangouski, M., & Nagasaka, M. (2000). Probing the binding pocket of the active site of aromatase with 6-ether or 6-ester substituted androst-4-ene-3,17-diones and their diene and triene analogs. *Steroids*, *65*(12), 871–882. doi:10.1016/S0039-128X(00)00169-0 PMID:11077085

Numazawa, M., Yamada, K., Watari, Y., & Ando, M. (2002). Improved synthesis and molecular modeling of 4β, 19-dihydroxyandrost-5-ene-17-one, an excellent inhibitor of aromatase. *Chemical & Pharmaceutical Bulletin*, *50*(5), 703–705. doi:10.1248/cpb.50.703 PMID:12036037

Okada, M., Yoden, T., Kawaminami, E., Shimada, Y., Kudoh, M., & Isomura, Y. (1997a). Studies on aromatase inhibitors. II. Synthesis and biological evaluation of 1-amino-1H-1, 2, 4-triazole derivatives. *Chemical & Pharmaceutical Bulletin*, *45*(2), 333–337. doi:10.1248/cpb.45.333 PMID:9118448

Okada, M., Yoden, T., Kawaminami, E., Shimada, Y., Kudoh, M., & Isomura, Y. (1997b). Studies on aromatase inhibitors. III. Synthesis and biological evaluation of [(4-bromobenzyl) (4-cyanophenyl) amino]azole and their azine analogs. *Chemical & Pharmaceutical Bulletin*, *45*(3), 482–486. doi:10.1248/cpb.45.482 PMID:9085555

Okada, M., Yoden, T., Kawaminami, E., Shimada, Y., Kudoh, M., & Isomura, Y. (1997c). Studies on aromatase inhibitors. IV. Synthesis and biological evaluation of N, N-disubstituted-5-aminopyrimidine derivatives. *Chemical & Pharmaceutical Bulletin*, *45*(8), 1293–1299. doi:10.1248/cpb.45.1293 PMID:9301028

Okada, M., Yoden, T., Kawaminami, E., Shimada, Y., Kudoh, M., & Isomura, Y. et al. (1996). Studies on aromatase inhibitors. I. Synthesis and biological evaluation of 4-amino-4H-1, 2, 4-triazole derivatives. *Chemical & Pharmaceutical Bulletin*, *44*(10), 1871–1879. doi:10.1248/cpb.44.1871 PMID:8904814

Ortuso, F., Langer, T., & Alcaro, S. (2006). GBPM: GRID based pharmacophore model, concept and application studies to protein-protein recognition. *Bioinformatics (Oxford, England)*, *22*(12), 1449–1455. doi:10.1093/bioinformatics/btl115 PMID:16567363

Osborne, C., & Tripathy, D. (2005). Aromatase inhibitors: Rationale and use in breast cancer. *Annual Review of Medicine*, *56*(1), 103–116. doi:10.1146/annurev.med.56.062804.103324 PMID:15660504

Osborne, C. K., & Schiff, R. (2005). Aromatase inhibitors: Future directions. *The Journal of Steroid Biochemistry and Molecular Biology*, *95*(1-5), 183–187. doi:10.1016/j.jsbmb.2005.04.021 PMID:16019204

Pasqualini, J. R. (2004). The selective estrogen enzyme modulators in breast cancer: A review. *Biochimica et Biophysica Acta*, *1654*(2), 123–143. PMID:15172700

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Pasqualini, J. R., Chetrite, G., Blacker, C., Feinstein, M. C., Delalonde, L., Talbi, M., & Maloche, C. (1996). Concentrations of estrone, estradiol, and estrone sulfate and evaluation of sulfatase and aromatase activities in pre- and postmenopausal breast cancer patients. *The Journal of Clinical Endocrinology and Metabolism*, *81*(4), 1460–1464. PMID:8636351

Petkov, P. I., Temelkov, S., Villeneuve, D. L., Ankley, G. T., & Mekenyan, O. G. (2009). Mechanism-based categorization of aromatase inhibitors: A potential discovery and screening tool. *SAR and QSAR in Environmental Research*, *20*(7–8), 657–678. doi:10.1080/10629360903438347 PMID:20024803

Piccart-Gebhart, M. J., Procter, M., Leyland-Jones, B., Goldhirsch, A., Untch, M., & Smith, I. et al. (2005). Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *The New England Journal of Medicine*, *353*(16), 1659–1672. doi:10.1056/NEJMoa052306 PMID:16236737

Plourde, P. V., Dyroff, M., & Dukes, M. (1994). Arimdex: A potent and selective fourth generation aromatase inhibitor. *Breast Cancer Research and Treatment*, *30*(1), 103–111. doi:10.1007/BF00682745 PMID:7949201

Polanski, J., & Gieleciak, R. (2003). The comparative molecular surface analysis (CoMSA) with modified uniformative variable elimination-PLS (UVE-PLS) Method: Application to the steroids binding the aromatase enzyme. *Journal of Chemical Information and Computer Sciences*, *43*(2), 656–666. doi:10.1021/ci020038q PMID:12653535

Pouget, C., Fagnere, C., Basly, J. P., Besson, A. E., Champavier, Y., Habrioux, G., & Chulia, A. J. (2002c). Synthesis and aromatase inhibitory activity of flavanones. *Pharmaceutical Research*, *19*(3), 286–291. doi:10.1023/A:1014490817731 PMID:11934235

Pouget, C., Fagnere, C., Basly, J. P., Habrioux, G., & Chulia, A. J. (2002a). Design, synthesis and evaluation of 4-imidazolylflavans as new leads for aromatase inhibition. *Bioorganic & Medicinal Chemistry Letters*, *12*(20), 2859–2861. doi:10.1016/S0960-894X(02)00565-6 PMID:12270163

Pouget, C., Fagnere, C., Basly, J. P., Habrioux, G., & Chulia, A. J. (2002b). New aromatase inhibitors. synthesis and inhibitory activity of pyridinyl-substituted flavanone derivatives. *Bioorganic & Medicinal Chemistry Letters*, *12*(7), 1059–1061. doi:10.1016/S0960-894X(02)00072-0 PMID:11909717

Pouget, C., Yahiaoui, S., Fagnere, C., Habrioux, G., & Chulia, A. J. (2004). Synthesis and biological evaluation of 4-imidazolylflavans as nonsteroidal aromatase inhibitors. *Bioorganic Chemistry*, *32*(6), 494–503. doi:10.1016/j.bioorg.2004.06.008 PMID:15530990

Proteau, P. J. (2011). Steroid hormones and therapeutically related compounds. In Wilson and Gisvold's textbook of organic medicinal and pharmaceutical chemistry (pp. 819-867). Philadelphia: Lipincott Williams and Wilkins.

Recanatini, M. (1996). Comparative molecular field analysis of non-steroidal aromatase inhibitors related to fadrozole. *Journal of Computer-Aided Molecular Design*, *10*(1), 74–82. doi:10.1007/BF00124467 PMID:8786417

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Recanatini, M., Bisi, A., Cavalli, A., Belluti, F., Gobbi, S., & Rampa, A. et al. (2001). A new class of nonsteroidal aromatase inhibitors: Design and synthesis of chromone and xanthone derivatives and inhibition of the P450 enzymes aromatase and 17r-hydroxylase/C17, 20-lyase. *Journal of Medicinal Chemistry*, *44*(5), 672–680. doi:10.1021/jm000955s PMID:11262078

Recanatini, M., & Cavalli, A. (1998). Comparative molecular field analysis of non-steroidal aromatase inhibitors: An extended model for two different structural classes. *Bioorganic & Medicinal Chemistry*, *6*(4), 377–388. doi:10.1016/S0968-0896(97)10053-0 PMID:9597182

Recanatini, M., Cavalli, A., & Valenti, P. (2002). Nonsteroidal aromatase inhibitors: Recent advances. *Medicinal Research Reviews*, *22*(3), 282–304. doi:10.1002/med.10010 PMID:11933021

Reed, M. J., Owen, A. M., Lai, L. C., Coldham, N. G., Ghilchik, M. W., Shaikh, N. A., & James, V. H. (1989). In situ estrogen synthesis in normal breast and breast tumor tissues: Effect of treatment with 4-hydroxyandrostenedione. *International Journal of Cancer*, *44*(2), 233–237. doi:10.1002/ijc.2910440208 PMID:2759729

Rollinger, J. M., Hornick, A., Langer, T., Stuppner, H., & Prast, H. (2004). Acetylcholinesterase inhibitory activity of scopolin and scopolectin discovered by virtual screening of natural products. *Journal of Medicinal Chemistry*, *47*(25), 6248–6254. doi:10.1021/jm049655r PMID:15566295

Roy, P. P., & Roy, K. (2010a). Molecular docking and QSAR studies of aromatase inhibitor androstenedione derivatives. *The Journal of Pharmacy and Pharmacology*, *62*(12), 1717–1728. doi:10.1111/j.2042-7158.2010.01154.x PMID:21054398

Roy, P. P., & Roy, K. (2010b). Docking and 3D-QSAR studies of diverse classes of human aromatase (CYP19) inhibitors. *Journal of Molecular Modeling*, *16*(10), 1597–1616. doi:10.1007/s00894-010-0667-y PMID:20195667

Ryan, B. M., Konecny, G. E., Kahlert, S., Wang, H. J., Untch, M., & Meng, G. et al. (2006). Survivin expression in breast cancer predicts clinical outcome and is associated with HER2, VEGF, urokinase plasminogen activator and PAI-1. *Annals of Oncology*, *17*(4), 597–604. doi:10.1093/annonc/mdj121 PMID:16403812

Saberi, M. R., Shah, K., & Simons, C. (2005). Benzofuran- and furan-2-yl-(phenyl)-3-pyridylmethanols: Synthesis and inhibition of P450 aromatase. *Journal of Enzyme Inhibition and Medicinal Chemistry*, *20*(2), 135–141. doi:10.1080/14756360400015256 PMID:15968818

Saberi, M. R., Vinh, T. K., Yee, S. W., Griffiths, B. J., Evans, P. J., & Simons, C. (2006). Potent CYP19 (aromatase) 1-[(Benzofuran-2-yl)(phenylmethyl)pyridine, -imidazole, and -triazole inhibitors: Synthesis and biological evaluation. *Journal of Medicinal Chemistry*, *49*(3), 1016–1022. doi:10.1021/jm0508282 PMID:16451067

Santen, R. J., & Harvey, H. A. (1999). Use of aromatase inhibitors in breast carcinoma. *Endocrine-Related Cancer*, *6*(1), 75–92. doi:10.1677/erc.0.0060075 PMID:10732791

Schlotter, C. M., Vogt, U., Allgayer, H., & Brandt, B. (2008). Molecular targeted therapies for breast cancer treatment. *Breast Cancer Research*, *10*(4), 211. doi:10.1186/bcr2112 PMID:18671839

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Schuster, D., Laggner, C., Steindl, T. M., Palusczak, A., Hartmann, R. W., & Langer, T. (2006). Pharmacophore modeling and in silico screening for new P450 19 (aromatase) inhibitors. *Journal of Chemical Information and Modeling*, *46*(3), 1301–1311. doi:10.1021/ci050237k PMID:16711749

Schwarzel, W. C., Kruggel, W. G., & Brodie, H. J. (1973). Studies on the mechanism of estrogen biosynthesis. 8. The development of inhibitors of the enzyme system in human placenta. *Endocrinology*, *92*(3), 866–880. doi:10.1210/endo-92-3-866 PMID:4267111

Seralini, G. E., & Moslemi, S. (2001). Aromatase inhibitors: Past, present and future. *Molecular and Cellular Endocrinology*, *178*(1-2), 117–131. doi:10.1016/S0303-7207(01)00433-6 PMID:11403901

Serpentini, C. N., Moslemi, S., Yous, S., Park, C. H., Lesieur, D., Sourdaine, P., & Seralini, G. E. (2004). Synthesis and evaluation of benzoxazolinoic imidazoles and derivatives of non-steroidal aromatase inhibitors. *Journal of Enzyme Inhibition and Medicinal Chemistry*, *19*(2), 119–127. doi:10.1080/14756360410001667319 PMID:15449726

Shaikh, S. A., Jain, T., Sandhu, G., Latha, N., & Jayaram, B. (2007). From drug target to leads-sketching a physicochemical pathway for lead molecule design in silico. *Current Pharmaceutical Design*, *13*(34), 3454–3470. doi:10.2174/138161207782794220 PMID:18220783

Shimozawa, O., Sakaguchi, M., Ogawa, H., Harada, N., Mihara, K., & Omura, T. (1993). Core glycosylation of cytochrome P-450(arom). Evidence for localization of N terminus of microsomal cytochrome P-450 in the lumen. *The Journal of Biological Chemistry*, *268*(28), 21399–21402. PMID:8407981

Siegel, R., Naishadham, D., & Jemal, A. (2012). Cancer statistics, 2012. *CA: a Cancer Journal for Clinicians*, *62*(1), 10–29. doi:10.3322/caac.20138 PMID:22237781

Simpson, E. R., Clyne, C., Rubin, G., Boon, W. C., Robertson, K., & Britt, K. et al. (2002). Aromatase-a brief overview. *Annual Review of Physiology*, *64*(1), 93–127. doi:10.1146/annurev.physiol.64.081601.142703 PMID:11826265

Simpson, E. R., & Davis, S. R. (2001). Minireview: Aromatase and the regulation of estrogen biosynthesis-some new perspectives. *Endocrinology*, *142*(11), 4589–4594. PMID:11606422

Simpson, E. R., Mahandroo, M. S., Means, G. D., Kilgore, M. W., Hinshelwood, M. M., & Graham-Lorence, S. et al. (1994). Aromatase cytochrome P450, the enzyme responsible for estrogen biosynthesis. *Endocrine Reviews*, *15*(3), 342–355. PMID:8076586

Simpson, E. R., Mahendroo, M. S., Means, G. D., Kilgore, M. W., Corbin, C. J., & Maldelson, C. R. (1993). Tissue-specific promoters regulate aromatase cytochrome P450 expression. *The Journal of Steroid Biochemistry and Molecular Biology*, *44*(4-6), 321–330. doi:10.1016/0960-0760(93)90235-O PMID:8476746

Sitteri, P. K., & Thompson, E. A. (1975). Studies of human placental aromatase. *Journal of Steroid Biochemistry*, *6*(3-4), 317–322. doi:10.1016/0022-4731(75)90149-1 PMID:810626

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Slamon, D. J., Leyland-Jones, B., Shak, S., Fuchs, H., Paton, V., & Bajamonde, A. et al. (2001). Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *The New England Journal of Medicine*, *344*(11), 783–792. doi:10.1056/NEJM200103153441101 PMID:11248153

Smith, I. E., & Dowsett, M. (2003). Aromatase inhibitors in breast cancer. *The New England Journal of Medicine*, *348*(24), 2431–2442. doi:10.1056/NEJMra023246 PMID:12802030

Spinelli, G. P., Tomao, F., Miele, E., Pasciuti, G., Russillo, M., & Tomao, S. (2008). Aromatase inhibitors in advanced breast cancer. *Recenti Progressi in Medicina*, *99*(1), 34–38. PMID:18389871

Stefanachi, A., Favia, A. D., Nicolotti, O., Leonetti, F., Pisani, L., & Catto, M. et al. (2011). Design, synthesis, and biological evaluation of imidazolyl derivatives of 4,7-disubstituted coumarins as aromatase inhibitors selective over 17-α-hydroxylase/C17-20 lyase. *Journal of Medicinal Chemistry*, *54*(6), 1613–1625. doi:10.1021/jm101120u PMID:21341743

Steindl, T. M., Schuster, D., Laggner, C., & Langer, T. (2006). Parallel screening: A novel concept in pharmacophore modeling and virtual screening. *Journal of Chemical Information and Modeling*, *46*(5), 2146–2157. doi:10.1021/ci6002043 PMID:16995745

Su, B., Hackett, J. C., Diaz-Cruz, E. S., Kim, Y. W., & Brueggemeier, R. W. (2005). Lead optimization of 7-benzyloxy 2-(40-pyridylmethyl)thio isoflavone aromatase inhibitors. *Bioorganic & Medicinal Chemistry*, *13*(23), 6571–6577. doi:10.1016/j.bmc.2005.07.038 PMID:16125392

Sun, B., Hoshino, J., Jermihov, K., Marler, L., Pezzuto, J. M., Mesecar, A. D., & Cushman, M. (2010). Design, synthesis, and biological evaluation of resveratrol analogues as aromatase and quinone reductase 2 inhibitors for chemoprevention of cancer. *Bioorganic & Medicinal Chemistry*, *18*(14), 5352–5366. doi:10.1016/j.bmc.2010.05.042 PMID:20558073

Taylor, H. M., Jones, C. D., Davenport, J. D., Hirsch, K. S., Kress, T. J., & Weaver, D. (1987). Aromatase inhibition by 5-substituted pyrimidines and dihydropyrimidines. *Journal of Medicinal Chemistry*, *30*(8), 1359–1365. doi:10.1021/jm00391a016 PMID:3612685

Tintori, C., Corradi, V., Magnani, M., Manetti, F., & Botta, M. (2008). Targets looking for drugs: A multistep computational protocol for the development of structure based pharmacophores and their applications for hit discovery. *Journal of Chemical Information and Modeling*, *48*(11), 2166–2179. doi:10.1021/ci800105p PMID:18942779

Tripos Inc. SYBYL-X 2.0, 1699 South Hanley Road, St. Louis, MO. 63144. USA, 2012.

Trunet, P. F., Vreeland, F., Royce, C., Chaudri, H. A., Cooper, J., & Bhatnagar, A. (1997). Clinical uses of aromatase inhibitors in the treatment of advanced breast cancer. *The Journal of Steroid Biochemistry and Molecular Biology*, *61*(3-6), 241–245. doi:10.1016/S0960-0760(97)80018-0 PMID:9365196

Ulmschneider, S., Muller-Vieira, U., Klein, C. D., Antes, I., Lengauer, T., & Hartmann, R. W. (2005). Synthesis and evaluation of (pyridylmethylene)tetrahydronaphthalenes/-indanes and structurally modified derivatives: Potent and selective inhibitors of aldosterone synthase. *Journal of Medicinal Chemistry*, *48*(5), 1563–1575. doi:10.1021/jm0492397 PMID:15743198

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Vinh, T. K., Ahmadi, M., Delgado, P. O., Perez, S. F., Walters, H. M., & Smith, H. J. et al. (1999). 1-[(Benzofuran-2-yl)phenylmethyl]-triazoles and -tetrazoles - potent competitive inhibitors of aromatase. *Bioorganic & Medicinal Chemistry Letters*, *9*(14), 2105–2108. doi:10.1016/S0960-894X(99)00328-5 PMID:10450990

Vinh, T. K., Yee, S. W., Kirby, A. J., Nicholls, P. J., & Simons, C. (2001). 1-[(Benzofuran-2-yl) phenylmethyl] triazoles as steroidogenic inhibitors:synthesis and in vitro inhibition of placental CYP19 aromatase. *Anti-Cancer Drug Design, 16*(4-5), 217-25.

Vogel, C. L., Cobleigh, M. A., Tripathy, D., Gutheil, J. C., Harris, L. N., & Fehrenbacher, L. et al. (2002). Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer. *Journal of Clinical Oncology*, *20*(3), 719–726. doi:10.1200/JCO.20.3.719 PMID:11821453

Wächter, G. A., Hartmann, R. W., Sergejew, T., Grun, G. L., & Ledergerber, D. (1996). Tetrahydronaphthalenes: Influence of heterocyclic substituents on inhibition of steroidogenic enzymes P450 arom and P450 17. *Journal of Medicinal Chemistry*, *39*(4), 834–841. doi:10.1021/jm950377t PMID:8632407

Wei, D., Jiang, X., Zhou, L., Chen, J., Chen, Z., & He, C. et al. (2008). Discovery of multi-target inhibitors by combining molecular docking with common pharmacophore matching. *Journal of Medicinal Chemistry*, *51*(24), 7882–7888. doi:10.1021/jm8010096 PMID:19090779

Westley, R. L., & May, F. E. B. (2013). A twenty-first century cancer epidemic caused by obesity: The involvement of insulin, diabetes, and insulin-like growth factors. *International Journal of Endocrinology*, 632461. PMID:23983688

Whomsley, R. F., Fernandez, E., Nicholls, P. J., Smith, H. J., Lombardi, P., & Pestillini, V. (1993). Substituted 1-[(Benzofuran-2-yl) phenylmethyl] imidazoles as potent inhibitors of aromatase in vitro and in female rats in vivo. *The Journal of Steroid Biochemistry and Molecular Biology*, *44*(4-6), 675–676. doi:10.1016/0960-0760(93)90279-6 PMID:8476782

Wolber, G., & Langer, T. (2005). LigandScout: 3D pharmacophores derived from protein bound ligands and their use as virtual screening filters. *Journal of Chemical Information and Modeling*, *45*(1), 160–169. doi:10.1021/ci049885e PMID:15667141

Woo, L. W. L., Bubert, C., Sutcliffe, O. B., Smith, A., Chander, S. K., & Mahon, M. F. et al. (2007). Dual aromatase-steroid sulfatase inhibitors. *Journal of Medicinal Chemistry*, *50*(15), 3540–3560. doi:10.1021/jm061462b PMID:17580845

Woo, L. W. L., Jackson, T., Putey, A., Cozier, G., Leonard, P., & Acharya, K. R. et al. (2010). Highly potent first examples of dual aromatase-steroid sulfatase inhibitors based on a biphenyl template. *Journal of Medicinal Chemistry*, *53*(5), 2155–2170. doi:10.1021/jm901705h PMID:20148564

Woo, L. W. L., Sutcliffe, O. B., Bubert, C., Grasso, A., Chander, S. K., & Purohit, A. et al. (2003). First dual aromatase-steroid sulfatase inhibitors. *Journal of Medicinal Chemistry*, *46*(15), 3193–3196. doi:10.1021/jm034033b PMID:12852749

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Wood, P. M., Woo, L. W. L., Humphreys, A., Chander, S. K., Purohit, A., Reed, M. J., & Potter, B. V. L. (2005). A letrozole-based dual aromatase-sulphatase inhibitor with in vivo activity. *The Journal of Steroid Biochemistry and Molecular Biology*, *94*(1-3), 123–130. doi:10.1016/j.jsbmb.2004.12.028 PMID:15862957

Wood, P. M., Woo, L. W. L., Labrosse, J. R., Thomas, M. P., Mahon, M. F., & Chander, S. K. et al. (2010). Bicyclic derivatives of the potent dual aromatase-steroid sulfatase inhibitor 2-Bromo-4-{[(4-cyanophenyl) (4H-1, 2, 4-triazol-4-yl) amino] methyl} phenylsulfamate: Synthesis, SAR crystal structure, and in vitro and in vivo activities. *ChemMedChem*, *5*(9), 1577–1593. doi:10.1002/cmdc.201000203 PMID:20632362

Wood, P. M., Woo, L. W. L., Labrosse, J. R., Trusselle, M. N., Abbate, S., & Longhi, G. et al. (2008). Chiral aromatase and dual aromatase-steroid sulfatase inhibitors from the letrozole template: Synthesis, absolute configuration, and in vitro activity. *Journal of Medicinal Chemistry*, *51*(14), 4226–4238. doi:10.1021/jm800168s PMID:18590272

Worachartcheewan, A., Mandi, P., Prachayasittikul, V., Toropova, A. P., Toropov, A. A., & Nantasenamat, C. (2014). Large-scale QSAR study of aromatase inhibitors using SMILES-based descriptors. *Chemometrics and Intelligent Laboratory Systems*, *138*(1), 120–126. doi:10.1016/j.chemolab.2014.07.017

Yahiaoui, S., Pouget, C., Fagnere, C., Champavier, Y., Habriouxa, G., & Chulia, A. J. (2004). Synthesis and evaluation of 4-triazolylflavans as new aromatase inhibitors. *Bioorganic & Medicinal Chemistry Letters*, *14*(20), 5215–5218. doi:10.1016/j.bmcl.2004.07.090 PMID:15380230

Zhao, Y., Agarwal, V. R., Mendelson, C. R., & Simpson, E. R. (1997). Transcriptional regulation of CYP19 gene (aromatase) expression in adipose stromal cells in primary culture. *The Journal of Steroid Biochemistry and Molecular Biology*, *61*(3-6), 203–210. doi:10.1016/S0960-0760(97)80013-1 PMID:9365191

Zhou, C., Zhou, D., Esteban, J., Murai, J., Siiteri, P. K., Wilczynski, S., & Chen, S. (1996). Aromatase gene expression and its exon I usage in human breast tumors. Detection of aromatase messenger RNA by reverse transcription polymerase chain reaction (RT-PCR). *The Journal of Steroid Biochemistry and Molecular Biology*, *59*(2), 163–171. doi:10.1016/S0960-0760(96)00100-8 PMID:9010331

ADDITIONAL READING

Balunas, M. J., Su, B., Brueggemeier, R. W., & Kinghorn, A. D. (2008). Xanthones from the botanical dietary supplement mangosteen (*Garcinia mangostana*) with aromatase inhibitory activity. *Journal of Natural Products*, *71*(7), 1161–1166. doi:10.1021/np8000255 PMID:18558747

Balunas, M. J., Su, B., Landini, S., Brueggemeier, R. W., & Kinghorn, A. D. (2006). Interference by naturally occurring fatty acids in a noncellular enzyme-based aromatase bioassay. *Journal of Natural Products*, *69*(4), 700–703. doi:10.1021/np050513p PMID:16643058

Bhatnagar, A. S. (2007). The discovery and mechanism of action of letrozole. *Breast Cancer Research and Treatment*, *105*(1), 7–17. doi:10.1007/s10549-007-9696-3 PMID:17912633

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Boccardo, F., Rubagotti, A., Puntoni, M., Guglielmini, P., Amoroso, D., & Fini, A. et al. (2005). Switching to anastrozole versus continued tamoxifen treatment of early breast cancer: Preliminary results of the Italian Tamoxifen Anastrozole Trial. *Journal of Clinical Oncology*, *23*(22), 5138–5148. doi:10.1200/JCO.2005.04.120 PMID:16009955

Brodie, A. H., & Mouridsen, H. T. (2003). Applicability of the intratumor aromatase preclinical model to predict clinical trial results with endocrine therapy. *American Journal of Clinical Oncology*, *26*(4), S17–S26. doi:10.1097/00000421-200308001-00004 PMID:12902873

Caporuscio, F., Rastelli, G., Imbriano, C., & Del Rio, A. (2011). Structure-Based Design of Potent Aromatase Inhibitors by High-Throughput Docking. *Journal of Medicinal Chemistry*, *54*(12), 4006–4017. doi:10.1021/jm2000689 PMID:21604760

Coupez, B., & Lewis, R. A. (2006). Docking and scoring--theoretically easy, practically impossible? *Current Medicinal Chemistry*, *13*(25), 2995–3003. doi:10.2174/092986706778521797 PMID:17073642

Doiron, J., Soultan, A. H., Richard, R., Touré, M. M., Picot, N., & Richard, R. et al. (2011). Synthesis and structureeactivity relationship of 1- and 2-substituted-1,2,3-triazole letrozole-based analogues as aromatase inhibitors. *European Journal of Medicinal Chemistry*, *46*(9), 4010–4024. doi:10.1016/j.ejmech.2011.05.074 PMID:21703734

Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., & Mainz, D. T. et al. (2004). GLIDE: A new approach for rapid, accurate, docking and scoring method and assessment of docking accuracy. *Journal of Medicinal Chemistry*, *47*(7), 1739–1749. doi:10.1021/jm0306430 PMID:15027865

Gobbi, S., Hu, Q., Negri, M., Zimmer, C., Belluti, F., & Rampa, A. et al. (2013). Modulation of cytochromes P450 with xanthone-based molecules: From aromatase to aldosterone synthase and steroid 11β-hydroxylase inhibition. *Journal of Medicinal Chemistry*, *56*(4), 1723–1729. doi:10.1021/jm301844q PMID:23363058

Goss, P., & von Eichel, L. (2007). Summary of aromatase inhibitor trials: The past and future. *The Journal of Steroid Biochemistry and Molecular Biology*, *106*(1-5), 40–48. doi:10.1016/j.jsbmb.2007.05.023 PMID:17627816

Hu, Q., Jagusch, C., Hille, U. E., Haupenthal, J., & Hartmann, R. W. (2010). Replacement of Imidazolyl by Pyridyl in Biphenylmethylenes Results in Selective CYP17 and Dual CYP17/CYP11B1 Inhibitors for the Treatment of Prostate Cancer. *Journal of Medicinal Chemistry*, *53*(15), 5749–5758. doi:10.1021/jm100317b PMID:20684610

Hu, Q., Yin, L., & Hartmann, R. W. (2013). Selective Dual Inhibitors of CYP19 and CYP11B2: Targeting Cardiovascular Diseases Hiding in the Shadow of Breast Cancer. *Journal of Medicinal Chemistry*, *55*(16), 7080–7089. doi:10.1021/jm3004637 PMID:22861193

Hu, Q., Yin, L., Jagusch, C., Hille, U. E., & Hartmann, R. W. (2010). Isopropylidene Substitution Increases Activity and Selectivity of Biphenylmethylene 4-Pyridine Type CYP17 Inhibitors. *Journal of Medicinal Chemistry*, *53*(13), 5049–5053. doi:10.1021/jm100400a PMID:20550118

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Huey, R., Morris, G. M., Olson, A. J., & Goodsell, D. S. (2007). A semiempirical free energy force field with charge-based desolvation. *Journal of Computational Chemistry*, *28*(6), 1145–1152. doi:10.1002/jcc.20634 PMID:17274016

Jain, A. (2003). Surflex: Fully automatic flexible molecular docking using a molecular similarity-based search engine. *Journal of Medicinal Chemistry*, *46*(4), 499–511. doi:10.1021/jm020406h PMID:12570372

Jakesz, R., Greil, R., Gnant, M., Schmid, M., Kwasny, W., & Kubista, E. et al. (2007). Extended Adjuvant Therapy With Anastrozole Among Postmenopausal Breast Cancer Patients: Results From the Randomized Austrian Breast and Colorectal Cancer Study Group Trial 6a. *Journal of the National Cancer Institute*, *99*(24), 1845–1853. doi:10.1093/jnci/djm246 PMID:18073378

Kondratyuk, T. P., Park, E. J., Marler, L. E., Ahn, S., Yuan, Y., & Choi, Y. et al. (2011). Resveratrol derivatives as promising chemopreventive agents with improved potency and selectivity. *Molecular Nutrition & Food Research*, *55*(8), 1249–1265. doi:10.1002/mnfr.201100122 PMID:21714126

Krammer, A., Kirchhoff, P. D., Jiang, X., Venkatachalam, C. M., & Waldman, M. (2005). LigScore: A novel scoring function for predicting binding affinities. *Journal of Molecular Graphics & Modelling*, *23*(5), 395–407. doi:10.1016/j.jmgm.2004.11.007 PMID:15781182

Lu, W. J., Xu, C., Pei, Z., Mayhoub, A. S., Cushman, M., & Flockhart, D. A. (2012). The tamoxifen metabolite norendoxifen is a potent and selective inhibitor of aromatase (CYP19) and a potential lead compound for novel therapeutic agents. *Breast Cancer Research and Treatment*, *133*(1), 99–109. doi:10.1007/s10549-011-1699-4 PMID:21814747

Maiti, A., Reddy, P. V., Sturdy, M., Marler, L., Pegan, S. D., & Mesecar, A. D. et al. (2009). Synthesis of casimiroin and optimization of its quinone reductase 2 and aromatase inhibitory activities. *Journal of Medicinal Chemistry*, *52*(7), 1873–1884. doi:10.1021/jm801335z PMID:19265439

Mayhoub, A. S., Marler, L., Kondratyuk, T. P., Park, E. J., Pezzuto, J. M., & Cushman, M. (2012). Optimization of the aromatase inhibitory activities of pyridylthiazole analogues of resveratrol. *Bioorganic & Medicinal Chemistry*, *20*(7), 2427–2434. doi:10.1016/j.bmc.2012.01.047 PMID:22386564

Mayhoub, A. S., Marler, L., Kondratyuk, T. P., Park, E. J., Pezzuto, J. M., & Cushman, M. (2012). Optimizing thiadiazole analogues of resveratrol versus three chemopreventive targets. *Bioorganic & Medicinal Chemistry*, *20*(1), 510–520. doi:10.1016/j.bmc.2011.09.031 PMID:22115839

Mooij, W. T., & Verdonk, M. L. (2005). General and targeted statistical potentials for protein-ligand interactions. *Proteins*, *61*(2), 272–287. doi:10.1002/prot.20588 PMID:16106379

Oda, A., Tsuchida, K., Takakura, T., Yamaotsu, N., & Hirono, S. (2006). Comparison of consensus scoring strategies for evaluating computational models of protein-ligand complexes. *Journal of Chemical Information and Modeling*, *46*(1), 380–391. doi:10.1021/ci050283k PMID:16426072

Rarey, M., Kramer, B., Lengauer, T., & Klebe, G. (1996). A fast flexible docking method using an incremental construction algorithm. *Journal of Molecular Biology*, *261*(3), 470–489. doi:10.1006/jmbi.1996.0477 PMID:8780787

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Raub, S., Steffen, A., Kamper, A., & Marian, C. M. (2008). AIScore chemically diverse empirical scoring function employing quantum chemical binding energies of hydrogen-bonded complexes. *Journal of Chemical Information and Modeling*, *48*(7), 1492–1510. doi:10.1021/ci7004669 PMID:18597446

Stauffer, F., Furet, P., Floersheimer, A., & Lang, M. (2012). New aromatase inhibitors from the 3-pyridyl arylether and 1-aryl pyrrolo[2,3-c]pyridine series. *Bioorganic & Medicinal Chemistry Letters*, *22*(5), 1860–1863. doi:10.1016/j.bmcl.2012.01.076 PMID:22335894

Velec, H. F. G., Gohlke, H., & Klebe, G. (2005). DrugScore knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. *Journal of Medicinal Chemistry*, *48*(20), 6296–9303. doi:10.1021/jm050436v PMID:16190756

Voets, M., Antes, I., Scherer, C., Muller-Vieira, U., Biemel, K., & Barassin, C. et al. (2005). Heteroaryl-substituted naphthalenes and structurally modified derivatives: Selective inhibitors of CYP11B2 for the treatment of congestive heart failure and myocardial fibrosis. *Journal of Medicinal Chemistry*, *48*(21), 6632–6642. doi:10.1021/jm0503704 PMID:16220979

Voets, M., Antes, I., Scherer, C., Muller-Vieira, U., Biemel, K., Marchais-Oberwinkler, S., & Hartmann, R. W. (2006). Synthesis and evaluation of heteroaryl-substituted dihydronaphthalenes and indenes: Potent and selective inhibitors of aldosterone synthase (CYP11B2) for the treatment of congestive heart failure and myocardial fibrosis. *Journal of Medicinal Chemistry*, *49*(7), 2222–2231. doi:10.1021/jm060055x PMID:16570918

Wang, R., Lai, L., & Wang, S. (2002). Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *Journal of Computer-Aided Molecular Design*, *16*(1), 11–26. doi:10.1023/A:1016357811882 PMID:12197663

Wang, R., Shi, H. F., Zhao, J. F., He, Y. P., Zhang, H. B., & Liu, J. P. (2013). Design, synthesis and aromatase inhibitory activities of novel indole-imidazole derivatives. *Bioorganic & Medicinal Chemistry Letters*, *23*(6), 1760–1762. doi:10.1016/j.bmcl.2013.01.045 PMID:23403081

Woo, L. W. L., Wood, P. M., Bubert, C., Thomas, M. P., Purohit, A., & Potter, B. V. (2013). Synthesis and structure-activity relationship studies of derivatives of the dual aromatase-sulfatase inhibitor 4-{[(4-cyanophenyl)(4H-1,2,4-triazol-4-yl)amino]methyl}phenyl sulfamate. *ChemMedChem*, *8*(5), 779–799. doi:10.1002/cmdc.201300015 PMID:23495205

Yahiaoui, S., Fagnere, C., Pouget, C., Buxeraud, J., & Chulia, A. (2008). New 7,8-benzoflavanones as potent aromatase inhibitors: Synthesis and biological evaluation. *Bioorganic & Medicinal Chemistry*, *16*(3), 1474–1480. doi:10.1016/j.bmc.2007.10.057 PMID:18042388

Yang, C., Wang, R., & Wang, S. (2006). M-score: A knowledge-based potential scoring function accounting for protein atom mobility. *Journal of Medicinal Chemistry*, *49*(20), 5903–5911. doi:10.1021/jm050043w PMID:17004706

Yin, L., Hu, Q., & Hartmann, R. W. (2013). Tetrahydropyrroloquinolinone type dual inhibitors of aromatase/aldosterone synthase as a novel strategy for breast cancer patients with elevated cardiovascular risks. *Journal of Medicinal Chemistry*, *56*(2), 460–470. doi:10.1021/jm301408t PMID:23281812

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Yin, S., Biedermannova, L., Vondrasek, J., & Dokholyan, N. V. (2008). MedusaScore: An accurate force field-based scoring function for virtual drug screening. *Journal of Chemical Information and Modeling*, *48*(8), 1656–1662. doi:10.1021/ci8001167 PMID:18672869

Zhang, C., Liu, S., Zhu, Q., & Zhou, Y. (2005). A knowledge-based energy function for protein-ligand, protein-protein and protein-DNA complexes. *Journal of Medicinal Chemistry*, *48*(7), 2325–2335. doi:10.1021/jm049314d PMID:15801826

Zhao, X., Liu, X., Wang, Y., Chen, Z., Kang, L., & Zhang, H. et al. (2008). An improved PMF scoring function for universally predicting the interactions of a ligand with protein, DNA and RNA. *Journal of Chemical Information and Modeling*, *48*(7), 1438–1447. doi:10.1021/ci7004719 PMID:18553962

KEY TERMS AND DEFINITIONS

Aromatase: Aromatase is a multienzymatic complex which is mostly overexpressed in breast cancer tissue and may be responsible for estrogen production and proliferation of breast tumors.

CoMFA: Comparative molecular field analysis is a ligand based drug design approach to find important steric and electrostatic fields of the ligands crucial for the biological activity.

Competitive Inhibitors: Competitive inhibitors compete with the substrate through non-covalent binding to the enzyme active site and block the enzymatic action.

CoMSIA: Comparative molecular similarity analysis is similar like CoMFA. Along with the important steric and electrostatic fields of the ligands, hydrophobic parameter is determined that can influence the biological activity.

Docking: A structure-based drug design approach where receptor/enzyme/protein and the ligand interactions are determined.

Mechanism-Based Inhibitors: Mechanism-based inhibitors mimic the substrate and converted to a reactive intermediate by the enzyme that inactivate the enzymatic action.

Molecular Properties: Those are the functional as well as the physicochemical characteristics of a molecule that is supposed to be responsible for the biological activity of molecules.

Pharmacophore Mapping: Important structural and physicochemical parameters responsible for potential biological activity may be determined.

QSAR: Quantitative structure-activity relationship is a ligand-based drug design approach where the important structural and physicochemical properties of the ligands are correlated with the biological activity to determine the important features.

This work was previously published in Quantitative Structure-Activity Relationships in Drug Design, Predictive Toxicology, and Risk Assessment edited by Kunal Roy, pages 400-470, copyright year 2015 by Medical Information Science Reference (an imprint of IGI Global).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 5
A Bounded Health Information Technology System Design Approach to Support Community-Based Care Delivery

Liam Peyton
University of Ottawa, Canada

Aladdin Baarah
University of Ottawa, Canada

Jaspreet Bindra
University of Ottawa, Canada

Austin Chamney
University of Ottawa, Canada

Craig Kuziemsky
University of Ottawa, Canada

ABSTRACT

Health information technology (HIT) offers great potential for supporting healthcare delivery, particularly collaborative care delivery that is provided across multiple settings and providers. To date much of HIT design has focused on digitizing data or processes on a departmental or healthcare provider basis. However, this bounded approach has not scaled well for supporting community based care across disparate providers or settings because of the lack of boundaries (e.g. disprate data and processes) that exist in community based care. Cloud computing approaches that leverage mobile form applications for developing integrated HIT solutions have the potential to support collaborative healthcare delivery in the community. However, to date there is a shortage of methods that describe how to develop integrated cloud computing solutions to support community based care delivery. In particular there is a need for methods that identify how to incorporate boundaries into cloud computing systems design. This paper uses a three year case study of the design of the Palliative Care Information System (PAL-IS) to provide system design insight on cloud computing approaches that leverage mobile forms applications to support community care management.

DOI: 10.4018/978-1-5225-0549-5.ch005

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

1. INTRODUCTION

Governments continue to spend large amounts of money on the adoption of health information technology (HIT) (Borycki, Newsham & Bates, 2013; Coiera, Aarts & Kulikowski, 2012; Blumental, 2009) to support the electronic exchange of health information (Fontaine, Ross, Zink & Schilling, 2010), continuity of care (Steward, Koester, Collins, Maiorana & Myers, 2010), and collaborative care delivery where patients are seen across a variety of settings and providers (Ellingsen, Monteiro & Røeda, 2013). These spending initiatives are in response to studies such as the Institute of Medicine report published in 2011 that concluded that the present healthcare trajectory has become too complex and costly and that digital technology will be a key aspect of healthcare delivery (Institute of Medicine, 2011). HIT has been purported to be a solution that will better provide services in the context of a shrinking workforce and an increased need for services (Coiera, 2009; Bates & Gawande, 2003).

However, while the need for HIT to enhance healthcare delivery is well described it has been far more challenging to develop systems that bring about needed change while minimizing adverse impacts (Pare & Sicotte, 2010). HIT implementations frequently cause unintended consequences including communication issues, creation of new work or more work, and even adverse events such as medical errors (Harrison, Koppel, & Bar-Lev, 2007). Unintended consequences occur for several reasons including poor fit with clinical workflow, differences in needs between different user groups (i.e. clinicians and administrators) or the co-existence of manual and automated processes (Novak, Brooks, Gadd, Anders, & Lorenzi, 2012). Unintended consequences can occur because of a gap between how HIT is used in practice versus what was intended by design, referred to as the "design-reality" gap (Heeks, 2006). To close this gap we will require new perspectives on HIT design to meet specific requirements. More specifically, new approaches to healthcare delivery such as collaborative care in the community will require the need to consider the needs of multiple users who may have very different process and information needs. Therefore the development of new approaches for systems design may be needed rather than just re-configuring existing systems (Karsh, Weinger, Abbott & Wears, 2010).

While we have made reasonably good progress on the development of hospital based systems, such as electronic health record (EHR) systems, we have not made as much progress at developing systems to support care across multiple settings where patients move between hospital and the community (Campion, Edwards, Johnson, & Kaushal, 2013). Care delivery in chronic illness and domains of medicine such as palliative care is a longitudinal process where patients frequently move between settings. Supporting care delivery across these varied settings is complex because of the dispersed nature and sheer number of processes that are provided (Bodenheimer, Wagner, & Grumbach, 2002). It is essential that we effectively integrate processes and data exchange across different settings or adverse events can occur (Health Quality Ontario, 2013).

In order to provide effective and continuous community based care delivery we need to coordinate care across complex networks of providers and various data sources (Health Quality Ontario, 2013; Brown, Lewis Ellis, Stewart, Freeman, & Kasperski, 2009; Bodenheimer, 2008; Kripalani, LeFevre, Phillips, Williams, Basaviah, & Baker, 2007). Cloud- based computing approaches offer great potential for providing the infrastructure to support continuous and collaborative care delivery. Cloud computing approaches can be particularly useful for providing the integrated architecture needed to support community based care across different settings (Kaur, & Chana, 2014; Bahga & Madisetti, 2013). Cloud-based approaches for HIT design offer advantages over traditional HIT applications in that data

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

and functionality is available to multiple users and resources. Cloud computing provides flexibility and reduces resource burden on a single setting (Chauhan, & Kumar, 2013). While the benefits of cloud computing for healthcare delivery such as common access to services and data across varied settings have been well described, system design approaches for cloud computing systems are less well known. Cloud computing requires a different system design approach from hospital or clinic based HIT design in that it requires integrating very diverse and possibly conflicting data or processes from multiple settings. While design of hospital based systems such as EHR systems have provided good insight about how to design and evaluate HIT, a shortcoming with these systems is they are 'bounded', and have defined users and requirements (Bjørn & Østerlund, 2009) In contrast community based collaborative care delivery is often unbounded and therefore the requirements, and more specifically quite often the processes that the requirements are meant to support, may be vague, or even unknown (Bodenheimer, Wagner, Grumbach, 2002). An essential first step for designing cloud based community systems is to identify systems design approaches to bring an element of bounding to the design of cloud computing systems.

A commonality in all system design approaches is that they have a grounding point from which design requirements are derived. For example, participatory design emphasizes user needs (Spinuzzi, 2005) and contexts while usability engineering emphasizes the need for systems to be designed that are enjoyable, safe and usable (Kushniruk, Nohr, Jensen & Borycki, 2013). Given the need for community based care delivery to bridge multiple settings and processes we suggest approaching systems design from the perspective of interoperability. Cloud based approaches need to support multiple types of interoperability. First is front line interoperability. If a system causes too many front line usability issues related to data entry and retrieval than it will lead to workflow and other problems (Kushniruk, Nohr, Jensen & Borycki, 2013). Second is the need to consider interoperability from organizational levels and the integration of data and processes across different settings. A third type of interoperability is the need for a system to support different tasks, such as clinical and performance management tasks. As there is increased pressure to make healthcare delivery more accountable (Chukmaitov, Harless, Bazzoli, Carretta & Siangphoe, 2014), goal-models (Yu, Giorgini, Maiden & Mylopoulos, 2011) are being used to capture organizational requirements for care-process management and then to link them to metrics for evaluation.

Overall there is a need for studies that identify strategies for systems design of cloud computing systems to support community care delivery. In particular is the need for the design of cloud computing systems to support different types of interoperability (i.e. clinical and performance management).

This paper is a revised and extended version of a conference paper where we described a system design approach for care process management through the use of a case study using an information system for palliative care (PAL-IS) (Baarah, Kuziemsky, Chamney, Bindra & Peyton, 2014). In this paper, we use the same case study to describe and evaluate our cloud computing system design approach for supporting community based palliative care delivery. First, we describe our method in terms of goals, design strategy and two development phases for the PAL-IS system. Second, we describe the phase 1 systems design, implementation and evaluation, including the identification of multiple interoperability issues that arose from phase 1. We then describe our revised cloud computing design approach we used in the design of PAL-IS phase 2 in order to provide boundaries to address the issues from phase 1. We then summarize our findings by comparing and contrasting the integrated cloud computing architecture and mobile forms approach employed in phase 2with the results from the traditional web application approach taken in phase 1.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

2. METHOD

2.1. Goals of PAL-IS

PAL-IS was designed with three overall goals. First were palliative care goals to provide consistent and continuous palliative care across different settings and providers. Second were care delivery goals that focused on enabling data collection and reporting, such as Palliative Performance Scale (PPS) trending, to enable better decision making around symptoms and disease trajectory. Third were service delivery goals that focused on performance management tracking of education activities, travel to and from homecare visits and the collection and provision of data to support analytics on palliative care delivery including numbers of cancer/non-cancer patients, and sources of patients referrals to support decision making on care delivery.

2.2. System Design Strategy

To achieve the above goals, the overall system design strategy was to design a system that would support a diverse set of requirements while also being usable for the end users. As such, our approach was to use iterative design-oriented research supplemented with participatory design (Spinuzzi, 2005; Hevner, March, Park & Ram, 2004). Design science consists of an iterative cycle of design, implementation, and evaluation of outcomes, in order to develop constructs, models, or methods (Hevner, March, Park, & Ram, 2004). Design science provides structure for obtaining and evaluating system requirements. Given the complex, and at times vague, nature of the processes and data we were trying to digitize we also needed a user engaged methodology. Participatory design (PD) provides such a method as it is based on users participating in the design (Spinuzzi, 2005; Grönvall, & Kyng, 2013). A combined design science-PD method enabled us to systematically achieve two outcomes. First was to have a structured and iterative approach for systems design and evaluation as per the design science approach. Second, was to use PD to incorporate user context and needs in the design, implementation, and evaluation of PAL-IS. One of the early challenges we encountered was scope as there is such a wide array of services and patient populations (i.e. inpatient, outpatient, and hybrid) in palliative care that it is difficult to design a system that supports all contexts. To manage the scope we focused on a specific palliative care program of the palliative pain and symptom management consultation service (PPSMCS) involving palliative care patients being seen by a family health team (FHT) consisting of a family doctor, a family medicine resident and a palliative care nurse. The objective of the FHT project was to better incorporate family physicians in palliative care delivery of order to increase the capacity to delivery palliative care services in the community.

2.3. Development Phases of PAL-IS

PAL-IS was developed in two phases. Phase 1 was a complete palliative care electronic health record (EHR) system designed to support the above FHT project. It was designed as a traditional web application. As part of the PD process we actively engaged members of PPSMCS including nursing staff, administers, physicians, IT staff, and data entry clerks to understand the workflow and data requirements for how patients were registered, assessed and reassessed during clinical encounters. The FHT project was 2 years in duration and phase 1 PAL-IS was used from July 2010 until June 2012. Approximately

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

100 patients were tracked in the system during that time. When phase 1 was evaluated there were several interoperability issues that were identified. Based upon this new understanding of the requirements and interoperability challenges, in phase 2 we developed a prototype cloud-based PAL-IS system that leveraged mobile forms.

3. RESULTS

We present our results in two parts. First is our description and analysis of phase 1 (complete palliative EHR implemented as a traditional web application). We describe the set of objectives PAL-IS was designed to achieve and how we designed PAL-IS to meet the objectives. We then evaluate the phase 1 implementation and identify interoperability issues that arose from phase 1. In phase 1 we focused on individual requirements rather than what is needed at the collaborative community level and as a result encountered several interoperability issues both for clinical use and performance management. In phase 1 we over engineered the clinical aspects and under engineered the performances management aspects.

Second we describe our revised phase 2 PAL-IS prototype and how it addressed or mitigated the interoperability issues identified in phase 1. We also assess it using the same interoperability criteria as phase 1. In phase 2 we took a step back and used goal metrics to drive requirements which provided more structure to the design and reduce the clinical data that needed to be collected, which reduced interoperability issues. More significantly, we used a different application architecture which greatly simplified coding of PAL-IS. Mobile forms for data collection were directly mapped to a multi-dimensional metrics data mart for performance management through the use of a cloud-base service.

3.1. Phase 1

PAL-IS phase 1 was implemented in July 2010 as a complete palliative care electronic health record system that was accessible from any web browser and PC. PAL-IS provided on-line forms that were identical to the paper documents used by the palliative care team.

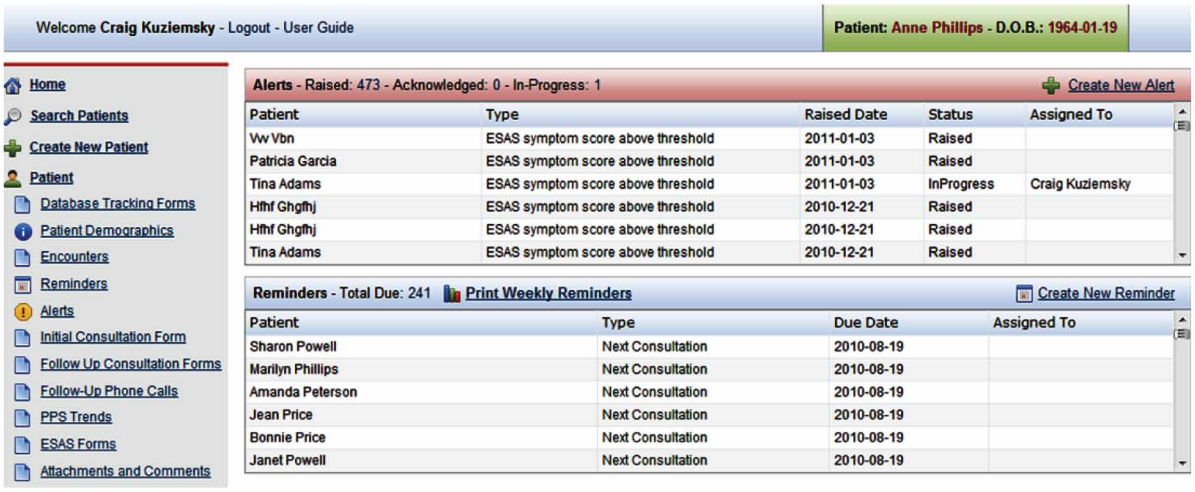
3.1.1. Phase 1 Systems Design

In terms of processes and information, PAL-IS focused on process support and data for consultations provided by palliative care team members. The consultation information maintained in PAL-IS included patient demographics, diagnoses, allergies, medications and contact information for family physician and any other providers (i.e. homecare providers) being seen by the patient. Supporting documents not originating in PAL-IS including lab test results, medical history, and other records from family physicians, specialist physicians (i.e. oncologists) and homecare providers could be scanned and added as attachments into PAL-IS. Similarly PAL-IS forms could be exported, printed and emailed as PDF documents that are identical to the paper forms used by the PPSMCS team.

Figure 1 shows the interface for the PAL-IS team portal that was implemented. On the left is a menu that provides access to all functions available in PAL-IS including forms and reports to document patient care, alerts, reminders and information about other members of the care team. The screen shot shows alerts and reminders for a particular patient.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 1. PAL-IS team portal



PAL-IS Phase 1 was successful in that it facilitated access to patient information regardless of location. It supported collaborative care delivery as PAL-IS clinicians were able to see what other members of the care team had done and could view patient symptoms over time to monitor and track trends. PAL-IS was also an integrated repository of data as it contained the electronic data about a patient but also attached documents to provide access to laboratory, diagnostic imaging or other clinical consultation reports.

3.1.2. Evaluation of Phase 1

PAL-IS Phase 1 was successful in that it successfully implemented the goals identified in the previous section. However from our meetings with end users we identified several interoperability issues over the two years that PAL-IS was used. We broadly categorize the interoperability issues into data, process and performance management issues.

Data interoperability was a big challenge because palliative care patients have many different care providers including a family physician, palliative care physician, and often a specialist physician such as an oncologist. Therefore, palliative care patients could have data in several electronic health (EHR) or medical record (EMR) systems. PAL-IS had no direct integration with these EHR or EMR systems. We investigated the potential for developing interfaces for the different systems but that task was not technically viable for several reasons. First, the sheer number of EMR systems in the community made it impractical. But more importantly, there was no standard interoperability engine for import/ export and a custom engine would have to be built specific to PAL-IS for each EMR.

Process interoperability issues refer to how PAL-IS impacted various palliative care processes. For example, one of the goals of PAL-IS was to support collaborative care delivery by providing alerts and reminders to ensure tasks were completed in a timely manner. However that goal was hard to achieve because of the manner in which data was entered and processed in PAL-IS. Front line usability issues meant that data was not always entered in real time, which had a ripple effect that impacted different PAL-IS functions. For example, the alerts and reminders system we developed required real time data

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

to be meaningful in order to ensure timely response to the issue raising the alert. However data was entered into PAL-IS in various ways (direct entry during consult, copy and paste from EMR, and retrospective entry after consult) and almost never in real-time. Real time data entry was one of the system requirements that arose from the PD sessions during the design of PAL-IS, However once PAL-IS had been implemented for several weeks it was apparent that 100% real time collection was not going to be plausible because the data entry burden was in conflict with clinical workflow. A further process issue was that PAL-IS was designed as a common interface and thus all users had to navigate through several screens and/or data fields while using PAL-IS. That required a significant information management effort and added to the poor workflow fit between PAL-IS and clinicians.

The most significant interoperability issue that we underestimated in phase 1 was that we did not understand that the most critical aspect of the system was the need to manage performance. During our PD requirements engineering meetings we focused mainly on data needs to support clinical practice such as patient consults and symptoms management. That approach greatly underestimated the interoperability that was needed in order to assess performance management. However, once the first version of PAL-IS was designed some of the administrators and senior physicians realized the value that electronic data could provide by enabling real time tracking of metrics required for governmental reporting. For example, several months into the implementation, a physician administrator wanted to track whether a resident was present during consults or whether the consult was an emergency visit. These data elements were quality improvement data that would support administrative decision making. However these data requirements were not identified until after implementation and the forms and the database needed to be redesigned to accommodate these new requests.

In summarizing the issues from phase 1, we designed PAL-IS without boundaries in order to solve immediate needs, some of which were excessive in terms of data collection and some of which were greatly underestimated at the start of the design process. The result of that was that we over-engineered the clinical aspects of PAL-IS and under-under engineered the performance management aspects.

3.2. Phase 2

PAL-IS phase 2 was implemented in June 2013 and was reinvented as a care process monitoring application (CPMA) that focused on maximizing the value of the performance management provided while minimizing the data entry burden. It was accessible from any web browser on a PC, tablet or phone in a much simpler, easier to use mobile forms interface customized for the different roles (clinician versus administrator). It collected the minimum data elements needed to drive reporting as a complement to the dictated notes that clinicians recorded and attached as transcribed notes.

In studying the interoperability issues from phase 1 we identified two key redesign requirements for phase 2. Foremost was that we needed to add boundaries to the systems design process to address the various interoperability issues that arose from the phase 1 evaluation. Two key design strategies allowed us to achieve that. First, we needed a better architectural separation and conceptual understanding of the relationship between data, processes and outcomes. PAL-IS phase 1 collected a substantial amount of data but over the course of its usage we realized that much of the data was not being used. Second, in order to address the process interoperability was that the interface for PAL-IS needed to be individualized for the different roles so that individuals only had to see or enter data specific to their needs.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

3.2.1. PAL-IS Phase 2 Systems Design

The phase 2 development methodology was driven by performance management requirements. Rather than focusing on solving immediate operational requirements as we did in phase 1, we drove our system design by analyzing the objectives of the care processes for the palliative care program and how these processes should be measured. The phase 2 system design had two steps. The first step was to develop an application model by analyzing the goals of the care process and defining how outcomes would be monitored using metrics. The goals of care and accompanying metrics became the boundaries for systems design. The second step was to develop a cloud-based application architecture that operationalized the application model, separating out the simple mobile forms interface from the metrics computed for reporting but maintaining the mapping between them. The two step method allowed us to map specific outputs from PAL-IS to the processes and people that would be using the outputs, thus enabled us to tailor the interfaces and output to specific process and users. Our hypothesis was that this would addresses the data, process, and technical interoperability issues from phase 1.

In the sections below we describe the application model, application architecture and the PAL-IS phase 2 application.

3.2.1.1. Application Model

The application model defines metrics to measure outcomes in terms of goals. Once metrics have been defined we determine what states of the care process these metrics correspond to and what data needs to be collected to compute the metrics. Forms are then designed and integrated into the flow of the care process to collect the required data. Table 1 shows an example of some of the metrics for PAL-IS phase 2 (there were 22 in total). For each metric, the form(s) that collect data for the metric is shown (e.g. "Referral" is the form used to collect date to measure "# Patients Cancer"). The first goal ("Quality of Care") is related to understanding the quality and range of care provided and is measured by calculating the relevant metrics as shown in the table. For example, a "Referral" form captures the data for "# Patients Cancer" (with drill down into diagnosis, gender, and age). To ensure wait time is minimized, the "Average Wait Time for Triage" (from "Referral" to "Appointment") is measured. The last quality of care metric ("# Patients Chemo Last 2 Weeks") is measured in order to track the number of unnecessary interventions during a patient's last days. The second goal ("Access") ensures adequate coverage of the patient population. For this goal, the number of referrals by physician and facility is captured to track patient referrals to palliative care across different facilities and providers. The "Education" goal is used to measure how effective the program is in promoting palliative care to physicians. To track this goal the "# of Consults With Resident Present" is tracked ("Consult" form).

Table 1. PAL-IS Phase 2 map of goals to metrics to forms

Goal	Metric	Form
Quality of Care	# Patients Cancer	Referral
	Average Wait Time for Triage	Referral, Appointment
	# Patients Chemo Last 2 Weeks	Deceased
Access	# Patients by Refer Facility	Referral
	# Patients by Refer Physician	Referral
Education	# Consults with Resident	Consult

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

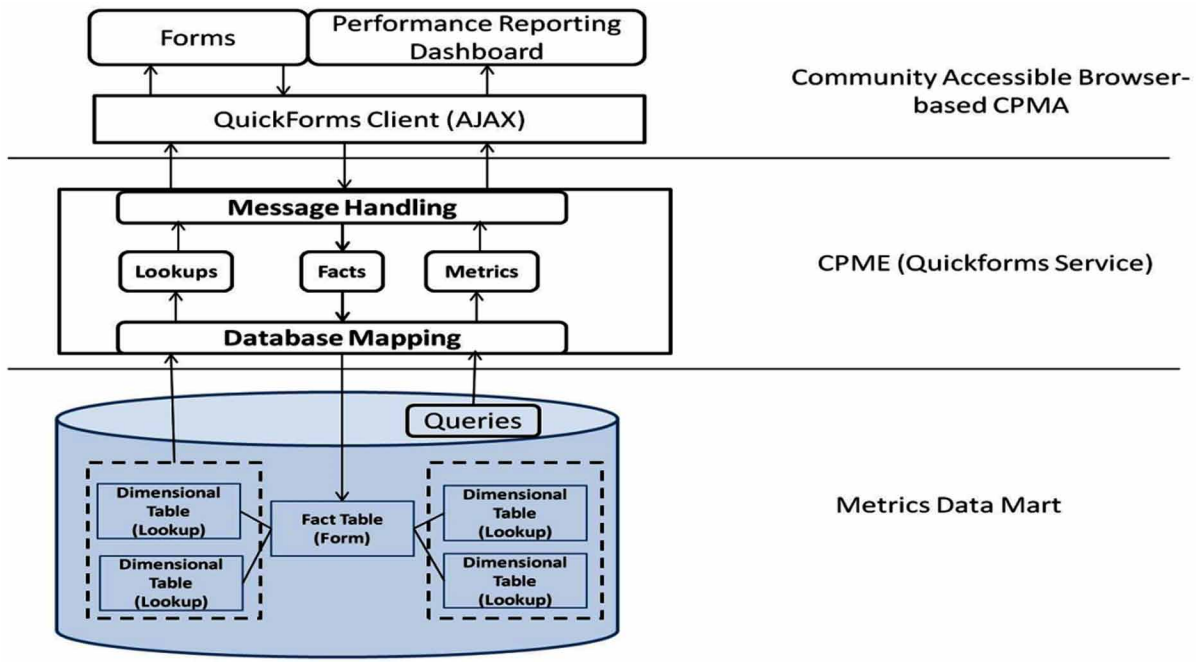
For each form the minimal dataset needed to be collected was determined in order to calculate all metrics. This helped to reduce the data and process interoperability burdens by only collecting data that is essential for monitoring the various goals.

3.2.1.2. Application Architecture

In order to simplify implementation of PAL-IS phase 2, we used a generic CPMA architecture, as shown in Figure 2, that mapped the mobile forms used for data collection to fact tables in a multi-dimensional metrics-data mart through a Care Process Management Engine (CPME) available in the cloud as a RESTful web service. In particular, the drop-down lists for each field of a data collection form were mapped to the dimensional lookup tables for the fact table. This simplified application development dramatically once the metrics data mart for reporting was defined. The data processing required by forms in their interaction with the database was performed automatically through an open source CPME called QuickForms. Figure 2 shows the application architecture. The CPMA provided mobile forms and performance management reports in HTML, supported by AJAX API calls in the QuickForms client to communicate over a REST interface to the QuickForms Service. The QuickForms Service provides database mapping to a multi-dimensional metrics data mart. Every time a form is filled in, it is automatically mapped and saves as a Fact in the data mart. Each of its fields is also mapped to dimensional lookup tables which are used to automatically populate the drop-down selection lists in the form. Finally, metrics are defined in terms of queries against the metrics data mart and displayed as reports in the Performance Reporting Dashboard.

The database is a classic star schema optimized for reporting. For each form, there is a fact table which stores the data collected by the form. Fields in forms have their own Dimension Tables that provide a lookup table of possible values for the field. The lookup values are generally classified into a hierarchy

Figure 2. PAL-IS Phase 2 application architecture

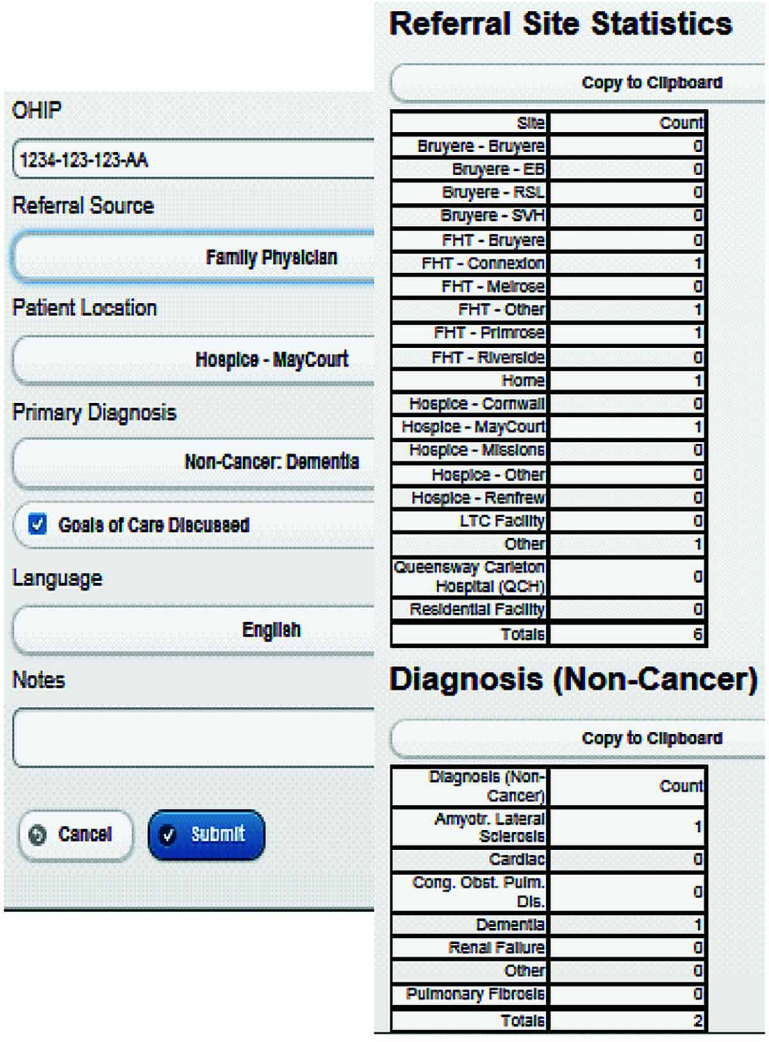


EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

to support drill down capabilities for reports. For example, a date field might be classified into day, month and year, so that one can see the number of referrals rolled up by day, month or year. The queries used to report metrics are stored in the database as well.

The Forms were built using jQuery Mobile to ensure compatibility across all device types and sizes (i.e. desktops, laptops, tablets, smartphones). The ability to use the application on multiple devices helped ensure the application was individualized and easy to use by all users in all roles. We tailored the report and form interfaces to correspond to specific roles. Unlike PAL-IS phase 1, where everyone saw a common interface and all forms and reports, the interface for phase 2 is tailored to only provide access to forms and reports that are relevant to a particular individual.

Figure 3. Referral form and reports



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

3.2.1.3. Application

Figure 3 shows a portion of the screenshots for the Referral form and its related reports. The user interface is much simpler and easier to use than the Phase 1 interface. Only the minimal elements needed to identify the patient (for reference by clinicians) and the attributes needed for performance management and care coordination are captured. Each reporting field is a simple drop down menu of precisely the categories that are used in reporting. This makes maintenance much easier as there is no need for developers to try and keep reports and forms in sync using the same labels. One set of labels is maintained in the database.

It also improves communication between clinicians and administrators. Clinicians have a better sense of what information is needed for performance management and how reports relate to their work. And there is a better attitude towards data collection, as the data entered is only the data that is considered critical and valuable for performance management or care coordination. Each role only sees the data they need to enter or use, and they understand why it is there.

3.2.2. Evaluation of PAL-IS Phase 2

We evaluated phase 2 using the same approach as phase 1. There were detailed walkthroughs, and interviews with users but with a greater focus on evaluating alignment between clinicians and administrators in terms of meeting all organizational requirements. Since the phase II implementation, PAL-IS had been positively reviewed by both the clinical and administrative team and was being used as a basis for building a regional palliative care registry and performance management dashboard. Initial feedback reveals that the revised design approach addressed or mitigated many of the issues from phase 1. PAL-IS Phase 2 is now being used as a basis for building a regional palliative care registry and performance management dashboard.

One major difference is that the data entry burden has been reduced by capturing only data needed to drive reporting. That was done to prevent the issue identified in the evaluation of phase 1 in that much of the data that was being entered (often double entered too) was not being used. The other key difference in phase 2 is we have made a substantial effort to reduce usability issues. The complexity of the forms and the amount of data entry for each form has been reduced.

There was also a huge improvement in terms of the efforts required to engineer PAL-IS. In particular, the design strategy of linking goals to metrics to forms ensured that requirements were clearly understood and encapsulated. It is an order of magnitude easier to maintain and update the application as requirements change. There is no longer a need for developers to make complex changes to general purposed programming code, as most changes are simply changes to data (lookup tables) in the database based on the application model or adjustments to the look of the form in HTML.

4. DISCUSSION

Cloud based systems offer great potential for supporting community based collaborative care delivery where patients move between multiple providers and settings. However, system design approaches for cloud computing systems differs from classic system design approaches in that defined boundaries need to be put around system design requirements.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 2 summarizes differences between Phase 1 and Phase 2. PAL-IS was designed and implemented in two phases. In phase 1 we took a classical systems design approach and built a traditional web application that digitized all existing paper forms that were used while employing several approaches (i.e. copy and pasting, scanning and attaching) for integrating other data. While phase 1 was successful, our longitudinal evaluation identified data, process and performance management interoperability issues. After evaluating phase 1 we realized our overall interoperability issue was that PAL-IS was unbounded and there were no limits on the data or processes we were trying to support. That resulted in a swelling of the data collection fields and processes it was trying to support. Designing PAL-IS without boundaries around data and processes resulted in a system that not meet the needs of either user group (clinical or administrative).

PAL-IS phase 2 was designed to address interoperability issues (i.e. process, data and performance management) from phase 1. Our approach for PAL-IS phase 2 was very different in that we started by developing an application model that defined goals and metrics for measuring the goals and used it to identify the minimal dataset of attributes that forms needed to collect.

As a result, requirements were more clearly defined and easier to manage and understand through a single application model. Even more importantly, we created an application architecture, optimized for care process monitoring applications (CPMA), that mapped forms directly to a multi-dimensional metrics data mart that was optimized for reporting. Because of this, the implementation effort was greatly reduced. It took one developer less than two months to develop phase 2, whereas phase 1 took two developers over a year. This was because phase 1 was entirely custom coded in a low level programming language (C#) while the vast majority of phase 2 was configurable, model-based forms and reports.

The advantage of our model and architecture (leveraging the QuickForms service) for cloud computing is it provides technical interoperability, scalability and reliability. The QuickForms service is a dedicated service for managing data collection and reporting in a hosted environment that provides mapping to and from forms or events to a multi-dimensional metrics data mart for reporting. This provides scalability both in terms of handling data collection across a community but also in terms of having the

Table 2. Comparison of PAL-IS Phase 1 and PAL-IS Phase 2

Criteria	Phase 1	Phase 2
Requirements	Ad hoc. Difficult. - “unbound”	Model-based. Clear - “Bound”
Architecture	Traditional web application architecture with database not optimized for reporting.	CPMA supported by cloud-based QuickForms service to map mobile forms to multi-dimensional metrics data mart.
Implementation Effort	Custom coded	Configurable, model-based forms and reports.
Data Interoperability - Entry Effort	50-100 fields per form. Complex fields.	~ 10 fields per form. Simple drop downs.
Data interoperability – External Systems	Manual (scanned PDFs from other systems).	QuickForms RESTful web service API defines data interchange.
Process Interoperability - Usability	Complex interface. Disconnected Reports	Simple custom interfaces Integrated reports.
Process Interoperability – Interface complexity	Common interface for all users requires excess information management effort	Interfaces individualized for different user roles
Performance Management Interoperability	Ad hoc, often requiring database or form redesign	Design-based integration based on defined goals and metrics

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

right infrastructure for reporting without requiring painful ETL processes. It is reliable because it does not have to be rewritten for every new care process monitoring application. It also addresses technical interoperability by defining the minimum data set in terms of facts for data collection that can be delivered either by mobile forms apps or by streamed events.

Data entry and usability was greatly simplified for users, both because there were fewer fields on the forms and also because they used drop down menus rather than typing. These small time saving features make a significant difference to providers who are often pressed for time while doing patient consults. Further, the interfaces were individualized to the different roles so users only saw what they needed to see. That design feature reduced the amount of information management that users had to do while interacting with PAL-IS.

Finally, performance management of palliative care was supported in a systematic, integrated fashion at design time rather than in an ad hoc manner after the fact. Because of the simplicity of the design, there is now a well-defined interface for data interchange with third party systems. This is enabling the next phase of PAL-IS as a regional palliative care registry and performance management dashboard.

One shortcoming to our approach is that bounding system requirements does require a trade-off between usability and functionality. While PAL-IS phase 1 was designed to support multiple functionalities it led to interoperability issues that impacted its usability. In Phase 2 goals and metrics were used to define boundaries, but that resulted in less process and data automation such as for patient consults. Future work will continue to study how design systems to achieve a proper balance between usability and functionality. A further limitation of our study is is that our findings and recommendations are based on a study of one system in one city. Different system design needs (i.e. goals and metrics) could emerge in other settings.

5. CONCLUSION

A significant challenge our healthcare system faces is the need to facilitate care delivery in a timely manner across multiple providers and settings. Cloud computing architectures can play a key role in facilitating community based collaborative care delivery but to date there is a lack of studies that have described system design approaches for community based cloud computing systems.

This paper provided a new bounded HIT system design strategy for to support community based collaborative care delivery that focuses on mapping organizational goals to metrics that measure performance based on a minimal dataset collected from forms. In particular, our approach is to build an application model of the HIT as a care process monitoring application (CPMA) and deliver it in an architecture that maps forms to metrics data marts optimized for reporting. Our approach is based on a three year case study of the design and implementation of a palliative care information system (PAL-IS).

The approach presented in this paper provides valuable insight for the design of HIT to support community based collaborative care delivery.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

REFERENCES

Baarah, A., Kuziemsky, C. E., Chamney, A., Bindra, J., & Peyton, L. (2014). A Design Strategy for Health Information Systems to Address Care Process Management, *proceedings of the 5th International Conference on Information and Communication Systems* doi:10.1109/IACS.2014.6841957

Bahga, A., & Madisetti, V. K. (2013). A Cloud-Based Approach to Interoperable Electronic Health Records (EHRs), *IEEE J. Biomedical and Health Informatics*, *17*(5), 894–906. doi:10.1109/JBHI.2013.2257818 PMID:25055368

Bates, D., & Gawande, A. (2003). Improving Safety with Information Technology. *The New England Journal of Medicine*, *348*(25), 2526–2534. doi:10.1056/NEJMsa020847 PMID:12815139

Bjørn, P., & Østerlund, C. (2009). Materiality in hospital work: Comparative study of the work in two pediatric emergency departments. In: AMCIS workshop paper, San Fransisco, USA.

Blumental, D. (2009). Stimulating the adoption of health information technology. *The New England Journal of Medicine*, *360*(15), 1477–1479. doi:10.1056/NEJMp0901592 PMID:19321856

Bodenheimer, T. (2008). Coordinating care—a perilous journey through the health care system. *The New England Journal of Medicine*, *358*(10), 1064–1071. doi:10.1056/NEJMhpr0706165 PMID:18322289

Bodenheimer, T., Wagner, E. H., & Grumbach, K. (2002). Improving primary care for patients with chronic illness. *Journal of the American Medical Association*, *288*(14), 1775–1779. doi:10.1001/jama.288.14.1775 PMID:12365965

Borycki, E.M., Newsham, D., & Bates, D.W. (2013). eHealth in North America. *IMIA Yearbook on Medical Informatics,* 103-106

Brown, J. B., Lewis, L., Ellis, K., Stewart, M., Freeman, T. R., & Kasperski, M. J. (2009). Mechanisms for communicating within primary health care teams. *Canadian Family Physician Medecin de Famille Canadien*, *55*(12), 1216–1222. PMID:20008604

Campion, T. R. Jr, Edwards, A. M., Johnson, S. B., & Kaushal, R. (2013). Health information exchange system usage patterns in three communities: Practice sites, users, patients, and data. *International Journal of Medical Informatics*, *82*(9), 810–820. doi:10.1016/j.ijmedinf.2013.05.001 PMID:23743323

Chauhan, R., & Kumar, A. (2013). Cloud computing for improved healthcare: Techniques, potential and challenges[EHB]. *E-Health and Bioengineering Conference*, *1*(4), 21–23.

Chukmaitov, A., Harless, D. W., Bazzoli, G. J., Carretta, H. J., & Siangphoe, U. (2014). Delivery system characteristics and their association with quality and costs of care: Implications for accountable care organizations.[epub ahead of print]. *Health Care Management Review*, 1. doi:10.1097/HMR.0000000000000014 PMID:24566250

Coiera, E. (2009). Building a National Health IT System from the Middle Out. *Journal of the American Medical Informatics Association*, *16*(3), 271–273. doi:10.1197/jamia.M3183 PMID:19407078

Coiera, E., Aarts, J., & Kulikowski, C. (2012). The dangerous decade. *Journal of the American Medical Informatics Association*, *0*(19), 2–5. doi:10.1136/amiajnl-2011-000674 PMID:22116642

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Ellingsen, G., Monteiro, E., & Røeda, K. (2013). Integration as interdependent workaround. *International Journal of Medical Informatics*, *82*(5), 161–169. doi:10.1016/j.ijmedinf.2012.09.004 PMID:23083928

Fontaine, P., Ross, S., Zink, T., & Schilling, L. M. (2010). Systematic Review of Health Information Exchange in Primary Care Practices. *Journal of the American Board of Family Medicine*, *23*(5), 655–670. doi:10.3122/jabfm.2010.05.090192 PMID:20823361

Grönvall, E., & Kyng, M. (2013). On participatory design of home-based healthcare. *Cognition Technology and Work*, *15*(4), 389–401. doi:10.1007/s10111-012-0226-7

Harrison, M. I., Koppel, R., & Bar-Lev, S. (2007). Unintended consequences of information technologies in health care: An interactive sociotechnical analysis. *Journal of the American Medical Informatics Association*, *14*(5), 542–549. doi:10.1197/jamia.M2384 PMID:17600093

Heeks, R. (2006). Health information systems: Failure, success and improvisation. *International Journal of Medical Informatics*, *75*(2), 125–137. doi:10.1016/j.ijmedinf.2005.07.024 PMID:16112893

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *Management Information Systems Quarterly*, *28*(1), 75–105.

Institute of Medicine. (2011). *Health IT and Patient Safety: Building Safer Systems for Better Care*. Washington: National Academics Press.

Karsh, B. T., Weinger, M. B., Abbott, P. A., & Wears, R. L. (2010). Health information technology: Fallacies and sober realities. *Journal of the American Medical Informatics Association*, *17*(6), 617–623. doi:10.1136/jamia.2010.005637 PMID:20962121

Kaur, P. D., & Chana, I. (2014). Cloud based intelligent system for delivering health care as a service. *Computer Methods and Programs in Biomedicine*, *113*(1), 346–359. doi:10.1016/j.cmpb.2013.09.013 PMID:24139021

Kripalani, S., LeFevre, F., Phillips, C. O., Williams, M. V., Basaviah, P., & Baker, D. W. (2007). Deficits in communication and information transfer between hospital-based and primary care physicians: Implications for patient safety and continuity of care. *Journal of the American Medical Association*, *297*(8), 831–841. doi:10.1001/jama.297.8.831 PMID:17327525

Kushniruk, A., Nohr, C., Jensen, S., & Borycki, E. M. (2013). From Usability Testing to Clinical Simulations: Bringing Context into the Design and Evaluation of Usable and Safe Health Information Technologies. *Yearbook of Medical Informatics*, 78–85. PMID:23974552

Novak, L., Brooks, J., Gadd, C., Anders, S., & Lorenzi, N. (2012). Mediating the intersections of organizational routines during the introduction of a health IT system. *European Journal of Information Systems*, *21*(5), 552–569. doi:10.1057/ejis.2012.2 PMID:24357898

Ontario, H. Q. (2013). Electronic tools for health information exchange: An evidence-based analysis. [Internet]. *Ontario Health Technology Assessment Series*, *13*(11), 1–76. Available from http://www.hqontario.ca/en/documents/eds/2013/full-report-OCDM-etools.pdf PMID:24194799

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Pare, G., & Sicotte, C. (2010). Success in health information exchange projects: Solving the implementation puzzle. *Social Science & Medicine*, *70*(8), 1159–1165. doi:10.1016/j.socscimed.2009.11.041 PMID:20137847

Spinuzzi, C. (2005). The Methodology of Participatory Design. *Technical Communication (Washington)*, *52*(2), 163–174.

Steward, W., Koester, K., Collins, S., Maiorana, A., & Myers, J. (2012). The essential role of reconfiguration capabilities in the implementation of HIV-related health information exchanges. *International Journal of Medical Informatics*, *81*(10), 10–20. doi:10.1016/j.ijmedinf.2012.07.004 PMID:22841703

Yu, E., Giorgini, P., Maiden, N., & Mylopoulos, J. (2011). *Social Modeling for Requirements Engineering: An Introduction*. MIT Press.

This work was previously published in the International Journal of Cloud Applications and Computing (IJCAC), 5(1); edited by Shadi Aljawarneh and Sam Goundar, pages 32-45, copyright year 2015 by IGI Publishing (an imprint of IGI Global).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 6
The Dark Side of Medical Tourism?
End of Life Choice, Human Trafficking, and Organ Transplants

Malcolm Cooper
Ritsumeikan Asia Pacific University, Japan

Mayumi Hieda
St Luke Clinic, Japan

ABSTRACT

There are 4 principles that should govern the response of the healthcare system in its treatment of individual medical problems. These may be summarized as: 1. medical care should be accessible to all; 2. the principle of patient autonomy should govern decision-making; 3. medical treatment should be recognized as being part of cultural behavior; and, 4. the medical profession should support the benefit of the patient. However, the combination of these with the rising cost of healthcare and the impact of globalization, has led to a dark side for medical tourism. In this situation, both patients and physicians are faced with ethical, human security and sustainability issues. This chapter examines 3 major issues in medical tourism: end of life choice, trafficking in human bodies and body parts, and organ transplants. In the healthcare systems of many countries, these issues can also involve criminal activities.

INTRODUCTION: TRENDS IN THE 'DARK' MEDICAL TOURISM MARKET

Traditionally, healthcare is disease oriented, rather than health oriented. Health orientated healthcare is focused mainly on behavior and lifestyle issues. In the medical model of disease, patients are usually not held responsible for the genesis of their illness. When a person becomes ill, the medical judgment implies that they cannot be blamed for their condition. This model goes on to specify that reactive treatment and care are appropriate and morally desirable (the medical treatment model). But the World Health Organization (WHO) has defined healthcare as "a process of enabling people to increase control

DOI: 10.4018/978-1-5225-0549-5.ch006

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

over and to improve their health." The strategy behind this notion is to relate personal choice and social responsibility in health to the creation of healthy environments, and to encourage positive personal health behaviors within these. However, current medical treatment arguments are not concerned with what individuals have done, but rather with how to repair the damage that resulted, and how they will behave in the future (Engelhardt, 2013).

A very great difficulty for the medical treatment model is that the financial burden of medical care expenses is increasing rapidly in many countries. Much of this rise in the cost of care can be attributed to advances in medical technology, but there is also a changing disease pattern at work too; 63% of all deaths worldwide currently stem from non-communicable diseases often requiring surgery (NCDs - chiefly cardiovascular diseases, cancers, chronic respiratory diseases, and diabetes; WHO, 2013). When these changing patterns conflict with local legal or healthcare structures, and shortages of the resources needed to solve some of them (replacement organs primarily), there rises the opportunity for criminal involvement and ethical disputes in the healthcare system (Scheper-Hughes, 2000; Denier, 2008; Neoh, 2014). Medical tourism can offset some of these issues, but paradoxically has contributed to some of them also. This chapter examines 3 major issues in medical tourism impacted by this situation that are being discussed at the time of writing: End of life choice, trafficking in human bodies and body parts, and organ transplants.

Dark Tourism

Arguably, despite our title, these issues do not absolutely conform to the original definition of dark tourism (also known as *thanatourism*): is the act of travel and visitation to sites, attractions and exhibitions that have real or recreated death, suffering or the seemingly macabre as a main theme (Stone, 2012; www.dark-tourism.org.uk). Tourist visits to former battlefields, slavery-heritage attractions, prisons, cemeteries, particular museum exhibitions, Holocaust sites, or to disaster locations all constitute the initial broad realm of 'dark tourism' (iTDR, 2014). However, this subject area has recently been receiving increasing attention from both academia and media alike (Seaton, 1999; Lennon & Foley, 2000; Miles, 2002; Ryan, 2005; Stone, 2006, 2012; IDTR, 2014). A growing body of interdisciplinary research has been undertaken that derives from the concept of '*death-related*' travel, and certainly, our first medical tourism issue of *End of Life* choice fits that definition (Jamal & Lelo, 2011), while human trafficking and organ transplants do also, when we consider the effects of such actions on donor communities and individuals. Thus, this chapter may also assist in our broader understanding of both the production and consumption of dark tourism, especially when we consider the relationships between this form of tourism and the cultural conditions and social institutions of contemporary societies relating to healthcare.

Stone (2012) has moved the conceptual base of dark tourism in a somewhat similar direction to where we think it should go in order to provide a theoretical context for dark medical tourism. Particularly, his work correlates dark tourism with questions of modern-day mortality (Stone & Sharpley, 2008). In this regard, researchers now often concentrate on the privatization and medicalization of death, which has become a relatively private experience (www.dark-tourism.org.uk). In relation to our discussion of medical tourism this secularization and privatization process has allowed medical treatment as technology to take over the dying process, and medical professionals and the State to deny choice as to the way to end life (Mellor & Shilling, 1993; Howarth, 2007; www.dark-tourism.org.uk). This approach also promotes attempts to maintain life through organ transplants and the like, that provide the opportunity for black markets in organs to develop (Scheper-Hughes, 2000). Our take on this situation is to expand the defini-

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

tion of dark tourism to include those people wishing to die that are kept alive by medical technology, and that are therefore forced to travel to seek a solution to their problem.

THE ISSUES

End of Life Choice

There are many reasons to include a discussion on the issues surrounding *end of life* treatment in an analysis of medical tourism, but perhaps the main one is that travel to another place is often required to obtain the treatment desired. This is because the medicalization of dying (Howarth, 2007; Stone, 2012; Hall, 2013) has been accompanied by the legal denial of end of life choice by the patient in many jurisdictions (Gawande, 2014). Often the best treatment a terminally ill patient can expect in this situation (before travel is factored in) is that palliative care will be available where medically assisted death is not allowed. The following list details the top ten reasons put forward (from Canadian sources) as to why end of life choice is necessary (Stone, 2012):

1. Palliative Care cannot cure all suffering. Palliative care is often critical, but alone it is not enough. In the US State of Oregon in 2012, 97% of individuals who used that state's dying with dignity legislation were registered for palliative care;
2. Control is a powerful pain reliever. Canadian studies have repeatedly shown that people are healthier when they know they have control over their lives (http://www.cehha.nshealth.ca/chb/people-healthy.htm). If all the Canadians who end their lives to avoid prolonged suffering could be assisted to die at a time of their choosing, they would not only have had far greater peace of mind, they may even have died a natural death;
3. Banning medically-assisted death does not in fact stop dying Canadians or anyone else from ending their lives. Many terminally-ill Canadians have travelled to Switzerland for example to receive this form of treatment, but the cost of this trip is beyond the means of most Canadians. Those who choose to end their lives in Canada are left to do so without medical assistance, which can lead to pain and suffering even if the attempt is successful. Others may choose to end their lives violently. This makes their death difficult not only for themselves, but also for their families and for emergency responders. Providing a peaceful means of death would prevent this trauma;
4. As a compassionate society, surely it is necessary to prevent needless suffering; offering medically assisted dying is compassionate. Family members and medical providers can be deeply traumatized when someone who is suffering greatly asks for relief when none can be provided;
5. Canadians in particular are already allowed to make choices that can end their lives, but are then denied the means to do so legally. By legalizing medically-assisted dying, the Country would be allowing all dying Canadians to have the same options as those who depend on machines for survival;
6. It is in fact the will of the majority of Canadians, and some Americans. A recent poll showed that 80% of Canadians support the right of grievously ill patients to have medical assistance to die;
7. Support for assisted dying has increased in jurisdictions where it has been legalized. Support for death with dignity has grown through the years in Oregon (USA), currently 80% of registered Oregon voters support the law, a growth in supporters of about 20% from the last time Oregon voters voted

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

on the law in 1997(http://www.deathwithdignity.org/2013/06/17/peg-sandeen-phd-msw). Once the choice is legalized, individuals realize fears are unwarranted and support for this option increases (Ganzini, Nelson, Lee, Kraemer, Schmidt, & Delorit, 2001);

8. Legalizing medically-assisted dying is associated with an improvement in end of life care for some patients. Jurisdictions that have legalized end-of-life choice have seen improvements in palliative care. For example, nurses who were asked to rate doctors on their palliative care skills in Oregon before and five years after legalization of medically assisted dying noted significant improvements (Goy, Jackson, Harvath, Miller, Delorit, M. A., & Ganzini, 2003);
9. Legalizing medically-assisted dying safeguards the weak and vulnerable. Right now, some medically assisted deaths happen without the patient's explicit consent. Studies in the Netherlands and Belgium have shown that the rate of these deaths dropped with the legalization of medically assisted dying. Post-legalization rates of deaths without explicit consent are also much lower in the Netherlands and in Belgium than in other jurisdictions studied (Australia and New Zealand), where assisted death is not legal (Mitchell & Owens, 2004).

The alternative view revolves around the nature of the physician's responsibilities (McMurray, Clarke, Barrasso, Clohan, Epps, Glasson, McQuillan, Plows, Puzak, Orentlicher, Halkola, &. Schweickart, 1992). Assisted end-of-life intervention occurs when a doctor facilitates a patient's death by providing the necessary means and/or information to enable the patient to perform the life-ending act (for example the physician provides sleeping pills and information about the lethal dose, while aware that the patient may commit suicide). However, allowing physicians to participate in assisted end-of-life may cause more harm than good. Physician-assisted suicide is fundamentally incompatible with the physician's role as healer, would be difficult or impossible to control, and would pose serious societal risks. Almost invariably, physicians must instead aggressively respond to the physical needs of patients at the end of life, and should not abandon them once it is determined that cure is impossible. While it is also emphasized that patients near the end of life must continue to receive emotional support, comfort care, adequate pain control, respect for patient autonomy, and good communication, these are generally subordinated to the need to keep them alive.

As a result, there is now considerable pressure to change this way of supporting end-of-life patients where they ask for this to be done (see for example the Manifesto of the Euthanasia Research & Guidance Organization (ERGO, 2009). This is to make medically hastened death by request lawful across jurisdictions as it is now in the Netherlands, Belgium, Luxembourg, Switzerland, Canada (Quebec, Ontario, British Columbia), and the American states of Oregon and Washington (each has different rules). It is legal to end your own life in in many countries, so making the extra step of help from doctors legal should not be an insurmountable problem. In turn this would place end-of-life medical tourism into the mainstream of alternative destinations for treatment based on the addition of the opportunity for *tourism* rather than as now, travel as the only alternative to restrictions back home.

Post-Script: The Maynard Case 29 October 2014

The American national debate about physician-assisted dying was brought to a head in late 2014, when a terminally ill patient ended her life by taking drugs available under an Oregon State law that allows such people to end their own lives (Maynard, 2014). The issue of physician-assisted dying is not new in America, but Maynard's youth and vitality before she became ill brought the matter to the attention

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

of the younger generation. This was assisted by the approach she took, to work with an NGO called *Compassion & Choices*, and to use her story to speak out for the right of terminally ill people to end their lives on their own terms. Her choice of course was not without detractors, some religious groups and others opposed to physician-assisted dying voiced objections.

Oregon was the first State of America to make it legal for a doctor to prescribe a life-ending drug to a terminally ill patient of sound mind who makes the request (1994): the patient must swallow the drug without help; it is in fact illegal for a doctor to administer it. Since the promulgation of this law more than 750 people, at an average age of 71, have used this facility to aid their decision-making. Oregon does not track how many terminally ill people move there to die; a patient must only prove to a doctor that they are living in the State. Only four other States in America – Washington, Montana, Vermont and New Mexico – allow patients to seek aid in dying.

Organ Transplants

The second of our issues, organ transplantation, is the best method of treatment for patients with end-stage organ failure (Sui, Yan, Xie, Chen, Li, Hu, Peng, & Dai, 2011; Sanwar Siraj, 2014). The following organs can be transplanted: heart, kidneys, liver, lungs, pancreas, intestine, and thymus (Al-Khader, 2002; WHO, 2008). And associated tissues include bones, tendons, cornea (eye parts), skin, heart valves, nerves and veins. On a global basis, kidneys are the most commonly transplanted organs, followed by the liver and heart. Eye and musculoskeletal grafts are the most commonly transplanted tissues, and are ten times more frequent than organ transplants. Kidneys and some other internal organs may be transplanted from living donors, but organs such as heart and lungs can only be obtained from brain dead patients. This is because these organs will be fresh from the oxygen supply received through ventilators, as brain dead patients are usually still on a life support machine (Al-Khader, 2002; WHO, 2008).

Transplantation raises a number of bioethical issues (Government of Bangladesh, 1999; Atighetchi, 2007), covered in Chapter 6 of this volume. These include definitions of death, when and how should consent be given and obtained for an organ to be transplanted, and the question of payment for organs (WHO, 2008). A particular problem that arises in this area of medicine is organ trafficking. The primary problem for domestic health care systems and medical tourism alike is the common shortage of organ donors. In Malaysia for example the rate of organ donation from deceased donors is among the lowest in the world. A report from the Malaysian *National Transplant Resource Centre* showed that there were only 26 actual donors in 2013. One of the causes of the small number of organ donors is the low rate of referral of brain dead patients. Referrals are made by doctors who treat patients diagnosed with brain death. Mortality due to brain death in intensive care units in some hospitals is high, but when brain death diagnosis is not done, then the patient would not be on record to be brain dead (Cohen, Ben Ami, Ashkenazi, & Singer, 2008; Sanwar Siraj, 2014).

When there is no diagnosis of brain death, then there will not be any referral. As a result, the number of potential donors is not be as high as it could be. However, this does not mean that if the referral rate is high then the number of donors is also high. At the end of the day, in Malaysia, whether a brain dead patient becomes a donor or not depends on the permission given by the next-of-kin (even if the patient is an organ pledger). Nonetheless, a high referral rate would show an effort among doctors, especially in the Intensive Care Unit, to diagnose brain death. Some of these patients diagnosed may turn out to be potential donors. The bottom line is if diagnosis of brain death is not done, then the probability of getting a donor is very low in the normal system. Thus, attitudes towards the acceptance of brain death

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

impact on the organ donation process. Most doctors who do not accept brain death are those who are not involved with the process of organ procurement or who are unfamiliar with the donation process. Abidin, Zainal, Wee Tong Ming, Loch, Hilmi, & Hautmann (2013) showed that there are still some doctors who are unwilling to accept the concept of brain death, although the Malaysian Medical Council (MMC) issued guidelines on the acceptance of brain death in 1996 (Merican, 2006; Hooi & Mansor, 2011). In this study, however, from the demographic data compiled through the questionnaires, none of the respondents had their medical background from countries which do not accept brain death as legal death.

On the other hand, individuals can donate their organs to family members or even strangers in most secular Western societies such as the UK. Two types of donation and transplantation are legally allowed in the UK: Directed and non-directed donation and transplantation (UKG, 2004, for England, Wales and Northern Ireland; UKG, 2006, for Scotland). The directed donation includes both genetically related kin (e.g. blood relatives like families) and emotionally related recipient (e.g. spouses, partners or close friends). Other types of directed donation, namely, paired and pooled are also allowed in the UK. When probable donors and recipients are mismatched, matches are allowed. At the same time, two or more donors and recipients are involved in this donation and transplantation practices. The second type of donation and transplantation is named *non-directed altruistic donation*. Altruistic donation is a non-directed donation where a donor is not clearly informed who the latent recipient will be, and is allowed in the UK, under the Human Tissue Authority's (HTA) approval.

Medical tourism is an alternative in this situation. Travel to receive an organ transplant, one of the medical miracles of the twentieth century now available in many destinations, has prolonged and improved the lives of hundreds of thousands of patients worldwide. Yet, in 2004 the World Health Organization called on member states: "*to take measures to protect the poorest and most vulnerable groups from transplant tourism and the sale of tissues and organs, including attention to the wider problem of international trafficking in human tissues and organs*" (WHO, 2004).

Human Body Part Trafficking

The demand for replacement organs has however led to trafficking in organs where demand outstrips supply. Our third issue for the growth of medical tourism is a particularly difficult one to address in single chapter. The growth of this industry has assisted patients from the outside world to seek treatment at a lower cost or within a shorter time period than at home (Smith, 2012), and boosted local, often developing economies. However, the expansion of the sector has been stigmatized by the growth of organ *trafficking* to facilitate transplantation to those who can afford it (Scheper-Hughes, 2000; Smith, 2012). Although developing countries have been taking measures to prohibit the sale of organs, the large gap between demand and supply has fuelled a black market that involves 'middlemen', medical personnel and the poor, as well as the intended recipient. This is particularly hard on the poor, whose abuse and exploitation is fuelled by the expansion of this trade, and the often illegitimate opportunities it creates (Scheper-Hughes, 2007). The problem is made worse by the low supply of organs in developed countries, where local people appear to be misinformed about the process, and hesitate to register as donors (Haustein & Sellers, 2004; Smith, 2012). The need for a national and international awareness campaigns is understood, as the expansion of medical tourism has the potential to encourage a further rise in organ trafficking (Connell, 2013; Smith, 2012; Hall, 2013).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Definitions

- Organ trafficking is the recruitment, transport, transfer, harboring or receipt of living or deceased persons or their organs by means of the threat or use of force or other forms of coercion, of abduction, of fraud, of deception, of the abuse of power or of a position of vulnerability, or of the giving to, or the receiving by, a third party of payments or benefits to achieve the transfer of control over the potential donor, for the purpose of exploitation by the removal of organs for transplantation (Siminoff, Arnold, & Hewlett, 2001; Declaration of Istanbul, 2008);
- Transplant commercialism is a policy or practice in which an organ is treated as a commodity, including by being bought or sold or used for material gain;
- Travel for transplantation is the movement of organs, donors, recipients or transplant professionals across jurisdictional borders for transplantation purposes. Travel for transplantation becomes transplant tourism if it involves organ trafficking and/or transplant commercialism, or if the resources (organs, professionals and transplant centers) devoted to providing transplants to patients from outside a country undermine the country's ability to provide transplant services for its own population (Turner, 2007).

The Increasing demand for donated organs, uncontrolled trafficking and the challenges of transplantation between closely-related species have prompted a serious re-evaluation of international guidelines and given new impetus to the role of WHO in gathering epidemiological data and setting basic normative standards (Medical News Today, 2004). While there are no reliable data on organ trafficking, it is widely believed to be on the increase, with brokers reportedly charging between US$100,000 and US$200,000 to organize a transplant for wealthy patients (Medical News Today, 2004; Gilbertson, Liu, Xue, Louis, Solid, Ebben, & Collins, 2005). The donors, frequently impoverished and ill-educated, may receive as little as US$1000 for a kidney although the going price is more likely to be about US$5000 (WHO, 2004).

A resolution adopted by the World Health Assembly (WHA) in 2002 voiced "concern at the growing insufficiency of available human material for transplantation to meet patient needs," and urged Member States to "extend the use of living kidney donations when possible, in addition to donations from deceased donors" (WHO, 2002; WHO, 2004; Medical News Today, 2004). It also urged governments "to take measures to protect the poorest and most vulnerable groups from 'transplant tourism' and the sale of tissues and organs, including attention to the wider problem of international trafficking in human tissues and organs." Since then, police have broken up an international ring which arranged for Israelis to receive kidneys from poor Brazilians at a clinic in the South African port city of Durban (Medical News Today, 2004), and countries such as Brazil, India, and Moldova, well-known sources of donors, have all banned the buying and selling of organs. Other countries are following suit. But this has come at the risk of driving the trade underground.

The growth in trafficking is based on an increasing demand for transplant organs. For example, one in 10 American adults, more than 20 million, have some level of chronic kidney disease (CKD), and the incidence of this disease has more than doubled since 2000 (NIH, 2012). Similarly, there is a high level of demand in Europe, where there are over 150,000 patients on dialysis treatment, and over 50,000 waiting for a kidney. The waiting list for a transplant in Europe has been between 7 and 10 years (2010). This is brought about by the guidelines on human organ transplantation adopted by the WHO in 1991. These state that organs should "be removed preferably from the bodies of deceased persons", and that live donors should in general be genetically related to the recipient (WHO, 2004). They also prohibit

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

"giving and receiving money, as well as any other commercial dealing" (Medical News Today, 2004). It should not be surprising therefore that a black market exists in organs, nor that medical tourists seek treatment in many less-than-ideal situations (Hall, 2013).

The Response of the Medical Industry

We have covered the responses of the medical industry to the question of more flexible and equitable end-of-life policies, in this section we address the industry's response to organ trafficking. The first question is that, even if trafficking related to the shortage of organs for transplantation could somehow be stopped, what would be the likely effects on *living* donors (Hippen & Taylor, 2007)? Without other means of achieving a transplant, medical tourists would need to rely on living donors even more than they currently do (Hippen & Taylor, 2007). The gender imbalance among living donors noted by Scheper-Hughes (2007) is real, and is worthy of further scrutiny (Hippen & Taylor, 2007). Also, international organ trafficking has generated demonstrable physical and psychological damage to vendors and recipients (Chugh & Jha, 1996, 2000; Ivanovski, Stojkovski, Cakalaroski, Masin, Djikova, & Polenakovic, 1997; Lawrence, 1997; Goyal, Mehta, Schneiderman, & Sehgal, 2002; Higgins, West, Fletcher, Stein, Lam, & Kashi, 2003; Jha, 2004), and is a real problem for reputable transplantation. The projected increase in the number of prevalent dialysis patients, and therefore the number of potential transplant recipients in the next few years (Gilbertson, Liu, Xue, Louis, Solid, Ebben, & Collins, 2005; Hippen, 2006a; Hippen & Taylor, 2007), concomitant with a comparatively stagnant growth in the number of deceased donor organs, will only increase pressure on living donors to offset the tourism brought about by inadequate public policy (Hippen, 2006b; Hippen & Taylor, 2007). These are the circumstances which threaten to render particularly vulnerable living donors even more so. These are problems that urgently warrant moral attention, which is why it is crucial, indeed, morally obligatory, that the challenges they pose for medical tourism are *accurately* explained and scrutinized (Hippen & Taylor, 2007).

CONCLUSION

There are many reasons to include a discussion on the issues surrounding end of life, trafficking in human bodies and body parts, and organ transplants as legitimate treatments in an analysis of the issues in medical tourism in the 21st Century, but perhaps the main one is that travel to another place may be required to obtain the treatment desired. Over the past decade or so, a growing body of interdisciplinary research has been undertaken that revolves around the concept of '*death-related*' travel, and certainly, these medical tourism issues fit that definition when we consider the effects of patient actions on donor communities and individuals. Thus, this chapter assists in our broader understanding of both the production and consumption of medical tourism, especially when we consider the relationships between this form of tourism, and the cultural conditions and social institutions of contemporary societies relating to healthcare.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

REFERENCES

Abidin, Z., & Zainal, L. (2013). Are health professionals responsible for the shortage of organs from deceased donors in Malaysia? *Transplant International*, *26*(2), 187–194. doi:10.1111/tri.12019 PMID:23199156

Al-Khader, A. A. (2002). The Iranian transplant program: comment from an Islamic perspective. *Nephrology, Dialysis, Transplantation*, *17*(2), 213–215. doi:10.1093/ndt/17.2.213 PMID:11812868

Atighetchi, D. (2007). *Islamic Bioethics: Problems and Perspectives*. Amsterdam: Springer. doi:10.1007/978-1-4020-4962-0

Chugh, K. S., & Jha, V. (1996). Commerce in transplantation in Third World countries. *Kidney International*, *49*(5), 1181–1186. doi:10.1038/ki.1996.171 PMID:8731080

Chugh, K. S., & Jha, V. (2000). Problems and outcomes of living unrelated donor transplants in the developing countries. *Kidney International*, *57*(s74), 131–135. doi:10.1046/j.1523-1755.2000.07421.x

Cohen, J., Ben Ami, S., Ashkenazi, T., & Singer, P. (2008). Attitude of Health Care Professionals to Brain Death: Influence on the Organ Donation Process. *Clinical Transplantation*, *22*(2), 211–215. doi:10.1111/j.1399-0012.2007.00776.x PMID:18339141

Connell, J. (2013). Contemporary Medical Tourism: Conceptualization, Culture and Commodification. *Tourism Management*, *34*, 1–13. doi:10.1016/j.tourman.2012.05.009

Declaration of Istanbul. (2008). *Report of the Convention on Transplant Tourism and Organ Trafficking*. Retrieved from http://declarationofistanbul.org/shdexltendy/er

Denier, Y. (2008). Mind the gap! Three approaches to scarcity in health care. *Medicine, Health Care, and Philosophy*, *11*(1), 73–87. doi:10.1007/s11019-007-9073-3 PMID:17508268

Engelhardt, H. T. Jr. (2013). The Family: Crucial to and Divisive in Bioethics, *International Journal of Chinese & Comparative Philosophy of Medicine,* 113-127.

Euthanasia Research & Guidance Organization (ERGO). (2009). *Liberty and Death: A manifesto concerning an individual's right to choose to die*. Oregon: ERGO. Retrieved from http://www.assistedsuicide.org/liberty_and_death_manifesto_right_to_die.html

Ganzini, L., Nelson, H. D., Lee, M. A., Kraemer, D. F., Schmidt, T. A., & Delorit, M. A. (2001). Oregon Physicians' Attitudes about and experiences with End-of-Life Care since passage of the Oregon Death with Dignity Act. *Journal of the American Medical Association*, *285*(18), 2363–2369. doi:10.1001/jama.285.18.2363 PMID:11343484

Gawande, A. (2014). *Being Mortal: Medicine and What Matters in the End*. Canada: Doubleday.

Gilbertson, D. T., Liu, J., Xue, J. L., Louis, T. A., Solid, C. A., Ebben, J. P., & Collins, A. J. (2005). Projecting the number of patients with end-stage renal disease in the United States to the year 2015. *Journal of the American Society of Nephrology*, *16*(12), 3736–3741. doi:10.1681/ASN.2005010112 PMID:16267160

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Government of Bangladesh. (1999). *Organ donation and transplantation policy*. Retrieved 21 June 2014, from http://bdlaws.minlaw.gov.bd/bangla_pdf_part.php?id=828&vol=32&search=1999

Goy, E. R., Jackson, A., Harvath, T., Miller, L. L., Delorit, M. A., & Ganzini, L. (2003). *Oregon hospice nurses and social workers' assessment of physician progress in palliative care over the past 5 years*. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16594421

Goyal, M., Mehta, R. L., Schneiderman, L. J., & Sehgal, A. R. (2002). Economic and health consequences of selling a kidney in India. *Journal of the American Medical Association*, *288*(13), 1589–1593. doi:10.1001/jama.288.13.1589 PMID:12350189

Hall, C. M. (Ed.). (2013). *Medical Tourism: The Ethics, Regulation, and Marketing of Health Mobility*. London: Routledge.

Haustein, S. V., & Sellers, M. T. (2004). Factors associated with (un)willingness to be an organ donor: Importance of public exposure and knowledge. *Clinical Transplantation*, *18*(2), 193–200. doi:10.1046/j.1399-0012.2003.00155.x PMID:15016135

Higgins, R., West, N., Fletcher, S., Stein, A., Lam, F., & Kashi, H. (2003). Kidney transplantation in patients travelling from the UK to India or Pakistan. *Nephrology, Dialysis, Transplantation*, *18*(4), 851–852. doi:10.1093/ndt/gfg026 PMID:12637669

Hippen, B. (2006a). Preventive measures may not reduce the demand for kidney transplantation. There is reason to suppose this is not the case. *Kidney International*, *70*(3), 606–607. doi:10.1038/sj.ki.5001719 PMID:16871262

Hippen, B. E. (2006b). The Case for Kidney Markets. *New Atlantis (Washington, D.C.)*, *14*(Fall), 47–61.

Hippen, B. E., & Taylor, J. S. (2007). In Defence of Transplantation: A Reply to Nancy Scheper-Hughes. *American Journal of Transplantation*, *7*(7), 1695–1697. doi:10.1111/j.1600-6143.2007.01840.x PMID:17511762

Hooi, L. S., & Mansor, L. Y. (2011). *Sixth Report of the National Transplant Registry Malaysia 2009* (Kuala Lumpur, 2011). Retrieved from http://www.crc.gov.my/pdf/6th NTR report2009.pdf

Howarth, G. (2007). *Death and dying: A sociological introduction*. Cambridge: Polity.

Ivanovski, N., Stojkovski, L., Cakalaroski, K., Masin, G., Djikova, S., & Polenakovic, M. (1997). Renal transplantation from paid, unrelated donors in India—it is not only unethical, it is also medically unsafe. *Nephrology, Dialysis, Transplantation*, *12*(9), 2028–2029. doi:10.1093/oxfordjournals.ndt.a027786 PMID:9306369

Jamal, T., & Lelo, L. (2011). Exploring the conceptual and analytical framing of dark tourism: from darkness to intentionality. In R. Sharpley & P. R. Stone (Eds.), *Tourist experience: Contemporary perspectives* (pp. 29–42). Abingdon, Oxon: Routledge.

Jha, V. (2004). Paid transplants in India: The grim reality. *Nephrology, Dialysis, Transplantation*, *19*(3), 541–543. doi:10.1093/ndt/gfg576 PMID:14767006

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Lawrence, R. (1997). Abuse of live related kidney transplantation. *Nephrology, Dialysis, Transplantation*, *12*(9), 2028. doi:10.1093/oxfordjournals.ndt.a027785 PMID:9306368

Lennon, J. J., & Foley, M. (2000). *Dark Tourism*. London: Continuum.

Maynard, B. (2014, November 3). My Right to Death with Dignity at 29. *CNN Report*.

McMurray, R. J., Clarke, O. W., Barrasso, J. A., Clohan, D. B., Epps, C. H. Jr, & Glasson, J. et al. (1992). Decisions near the end of life. *Journal of the American Medical Association*, *267*(16), 2229–2233. doi:10.1001/jama.1992.03480160087040 PMID:1372944

Medical News Today. (2004, September 7). *Organ Trafficking and Transplantation Pose New Challenges.* Bexhill-On-Sea, UK: Medilexicon International Ltd.

Mellor, P. A., & Shilling, C. (1993). Modernity, Self-Identity and the Sequestration of Death. *Sociology*, *27*(3), 3411–3431. doi:10.1177/0038038593027003005

Merican, M. I. (2006). *Guideline of the Malaysia Medical Council*. Kuala Lumpur: Malaysia Medical Council.

Miles, W. F. S. (2002). Auschwitz: Museum Interpretation and Darker Tourism. *Annals of Tourism Research*, *29*(4), 1175–1178. doi:10.1016/S0160-7383(02)00054-3

Mitchell, K., & Owens, G. (2004). End of life decision-making by New Zealand general practitioners: A national survey. *The New Zealand Medical Journal*, *117*(1196), U934. PMID:15280938

National Institutes of Health (NHI). (2012). *Kidney Disease Statistics for the United States*. Washington: NHI.

Neoh, C.-A. (2014). The need for and ethical handling of East-West-Native integrated global medicine. In *Proceedings of the 15th Asian Bioethics Conference (ABC 15) & 12th Asia Pacific Conference (APC 12),* (pp. 348-352). Beppu: Ritsumeikan Asia Pacific University.

Ryan, C. (2005). Dark tourism – an introduction. In C. Ryan, S. Page, & M. Aitken (Eds.), *Taking Tourism to the Limits: Issues, Concepts and Managerial Perspectives* (pp. 187–190). Oxford: Elsevier. doi:10.1016/B978-0-08-044644-8.50020-5

Scheper-Hughes, N. (2000). The Global Traffic in Human Organs. *Current Anthropology*, *41*(2), 191–224. doi:10.1086/300123 PMID:10702141

Scheper-Hughes, N. (2007). The tyranny of the gift: Sacrificial violence in living donor transplants. *American Journal of Transplantation*, *7*(3), 507–511. doi:10.1111/j.1600-6143.2006.01679.x PMID:17250554

Seaton, A. V. (1999). War and Thanatourism: Waterloo 1815-1914. *Annals of Tourism Research*, *26*(1), 130–158. doi:10.1016/S0160-7383(98)00057-7

Siminoff, L. A., Arnold, R. M., & Hewlett, J. (2001). The process of organ donation and its effect on consent. *Clinical Transplantation*, *15*(1), 39–47. doi:10.1034/j.1399-0012.2001.150107.x PMID:11168314

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Siraj, S. (2014). Organ donation and family relations in Bangladesh. In *Proceedings of the 15th Asian Bioethics Conference (ABC 15) & 12th Asia Pacific Conference (APC 12),* (pp. 57-60). Beppu: Ritsumeikan Asia Pacific University.

Smith, A. (2012). Medical Tourism and organ trafficking. *African Journal of Hospitality*. *Tourism and Leisure*, *2*(1), 1–9.

Stone, P. (2014). *Frequently Asked Questions*. The Institute for Dark Tourism Research (ITDR). Retrieved from www.dark-tourism.org.uk/faqs

Stone, P. R. (2006). A dark tourism spectrum: Towards a typology of death and macabre related tourist sites, attractions and exhibitions. *TOURISM*. *Special Issue on Spirituality and Meaningful Experiences in Tourism*, *52*(2), 145–160.

Stone, P. R. (2012). Dark Tourism and Significant Other Death: Towards a Model of Mortality Mediation. *Annals of Tourism Research*, *39*(3), 1565–1587. doi:10.1016/j.annals.2012.04.007

Stone, P. R., & Sharpley, R. (2008). Consuming Dark Tourism: A Thanatological Perspective. *Annals of Tourism Research*, *35*(2), 574–596. doi:10.1016/j.annals.2008.02.003

Sui, W. G., Yan, Q., Xie, S. P., Chen, H. Z., Li, D., & Hu, C. X. et al. (2011). Successful Organ Donation from Brain Dead Donors in a Chinese Organ Transplantation Center. *American Journal of Transplantation*, *11*(10), 2247–2249. doi:10.1111/j.1600-6143.2011.03694.x PMID:21883900

Turner, L. (2007). Medical Tourism: Family Medicine and International Health-Related Travel. *Canadian Family Physician Medecin de Famille Canadien*, *53*(10), 1639–1641. PMID:17934018

United Kingdom Government. (2004). *Human Tissue Act*. Retrieved from http://www.legislation.gov.uk/ukpga/2004/30/pdfs/ukpga_20040030_en.pdf

United Kingdom Government. (2006). *Human Tissue Act (Scotland)*. Retrieved from http://www.legislation.gov.uk/asp/2006/4/pdfs/asp_20060004_en.pdf

World Health Organization. (2002). *The world health report: reducing risks, promoting healthy life*. Geneva: World Health Organization.

World Health Organization. (2004, September 1). Organ Trafficking And Transplantation Pose New Challenges. Geneva: World Health Organization.

World Health Organization (WHO). (2008). *WHO Guiding Principles on human cell, tissue and organ transplantation*. Geneva: World Health Organization.

KEY TERMS AND DEFINITIONS

End of Life Choice: The medicalization of dying has been accompanied by the legal denial of end-of-life choice by the patient in many jurisdictions. Often the best treatment a terminally ill patient can expect in this situation (before travel is factored in) is that palliative care will be available where medically assisted death is not allowed. Thus travel to another place is often required to obtain the treatment desired.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Human Organ Trafficking: The demand for replacement organs has however led to trafficking in organs when demand outstrips supply. The growth of this industry has assisted medical tourists to seek treatment at a lower cost or within a shorter time period than at home, and has boosted local, often developing economies. However, the expansion of the sector has been stigmatized by the growth of organ *trafficking* to facilitate transplantation to those who can afford it. The large gap between demand and supply has fuelled a black market that involves 'middlemen', medical personnel and the poor, as well as the intended recipient.

Living Donors: Without other means of achieving a transplant, medical tourists are finding it increasingly necessary to rely on living donors, with a concomitant rise in legal and donor problems.

Organ Transplants: Is the best method of treatment for patients with end-stage organ failure. The following organs can be transplanted: heart, kidneys, liver, lungs, pancreas, intestine, and thymus, and associated tissues including bones, tendons, cornea (eye parts), skin, heart valves, nerves and veins can also. On a global basis, kidneys are the most commonly transplanted organs, followed by the liver and heart.

Thanatourism: Is the act of travel and visitation to sites, attractions and exhibitions that have real or recreated death, suffering or the seemingly macabre as a main theme. However, a growing body of interdisciplinary research has been undertaken that derives from the concept of '*death-related*' travel, and certainly the issue of End of Life choice fits that definition, while human trafficking and organ transplants do also, when we consider the effects of such actions on donor communities and individuals.

This work was previously published in Current Issues and Emerging Trends in Medical Tourism edited by Malcolm Cooper, Kazem Vafadari, and Mayumi Hieda, pages 232-244, copyright year 2015 by Medical Information Science Reference (an imprint of IGI Global).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 7
Red Meat and Health:
Evidence Regarding Red Meat, Health, and Chronic Disease Risk

Kate Marsh
Northside Nutrition and Dietetics, Australia

Angela Saunders
Sanitarium Health and Wellbeing, Australia

Carol Zeuschner
Sydney Adventist Hospital, Australia

ABSTRACT

Despite its nutritional benefits, there is an increasing body of evidence to suggest that regular consumption of red meat may negatively impact health and disease risk, including the risk of most common chronic diseases. This chapter reviews the current evidence linking red and processed meat intakes with chronic disease, obesity and mortality risks and discusses possible mechanisms to explain these associations. Research on the health benefits of diets low in red meat, including vegetarian, vegan, Mediterranean and other plant-based diets, is also reviewed.

INTRODUCTION

While red meat is a good source of protein and other nutrients including iron, zinc and vitamin B12, there is an increasing body of evidence to suggest that higher intakes of red meat may negatively impact health and disease risk.

Epidemiological studies have found a positive relationship between intakes of red meat and processed meats and disease risk including risk of obesity (Wang & Beydoun, 2009; R1osell, Appleby, Spencer & Key, 2006; Vergnaud et al., 2010; Rizzo, Jaceldo-Siegl, Sabate & Fraser, 2013), type 2 diabetes (Micha, Michas & Mozaffarian, 2012; Pan et al., 2011; Pan et al., 2013; Aune, Ursin & Veierod, 2009; Feskens, Sluik & van Woudenbergh, 2013), coronary heart disease (Micha et al., 2012; Clifton, 2011; Micha, Wallace & Mozaffarian, 2010; Bernstein et al., 2010), stroke (Chen, Lv, Pang & Liu, 2013; Kaluza, Wolk

DOI: 10.4018/978-1-5225-0549-5.ch007

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

& Larsson, 2012; Micha et al., 2010) and some types of cancer (particularly colorectal cancer) (Abid, Cross & Sinha, 2014), independent of other lifestyle factors. All-cause mortality, and cardiovascular and cancer mortality, are also increased with higher intakes of meat, and in particular processed meat (Larsson & Orsini, 2014; Sinha, Cross, Graubard, Leitzmann & Schatzkin, 2009; Pan et al., 2012; Rohrmann et al., 2013; Abete, Romaguera, Vieira, Lopez de Munain & Norat, 2014).On the other hand, plant-based diets which exclude or contain low intakes of red meat, including vegan, vegetarian, Mediterranean, Portfolio and DASH (Dietary Approaches to Stop Hypertension) diets, are associated with improved health outcomes (McEvoy, Temple & Woodside, 2012; Sievenpiper & Dworatzek, 2013).

There are a number of possible explanations for the association between intakes of meat and chronic disease risk. In particular, typical nutrients and other compounds present in red meat including saturated fatty acids, dietary cholesterol, animal protein, haem iron, Advanced Glycation End Products (AGE's), trimethylamine-N-oxide (TMAO), L-carnitine, Neu5Gc, nitrates and nitrosamines in processed and cooked meats are likely to play a role (Uribarri et al., 2010; Feskens et al., 2013).

This chapter reviews the evidence linking high intakes of red meat with chronic disease risk and mortality and discusses the possible mechanisms underlying these associations. It also explores the health benefits of dietary patterns low in red meat. With an increased interest in lower carbohydrate high protein and hunter-gatherer 'paleo' style diets, this is a timely review of the evidence into red meat and health.

BACKGROUND

Chronic diseases including heart disease, stroke, cancer and diabetes, are the leading causes of mortality, contributing to 60% of all deaths worldwide (WHO, 2014b). Cardiovascular disease (CVD) is the leading cause of death with an estimated 17.3 million people dying from CVD in 2008 (30% of all deaths worldwide) a figure that is expected to increase to 23.3 million by 2030 (WHO, 2014b). Of these deaths, the majority were due to coronary heart disease (CHD) (7.3 million) and stroke (6.2 million). Diabetes currently affects an estimated 347 million people worldwide and in 2012 was the direct cause of 1.5 million deaths (WHO, 2014c). The World Health Organisation (WHO) projects that diabetes will be the 7th leading cause of death in 2030. Cancer is also a leading cause of death worldwide, accounting for 8.2 million deaths in 2012, with lung, liver, stomach, colorectal and breast cancers causing the most cancer deaths each year (WHO, 2014a). The WHO estimates that annual cancer cases will rise from 14 million in 2012 to 22 million within the next two decades (WHO, 2014a).

Lifestyle factors, particularly diet, play a significant role in the risk of chronic diseases and diet related chronic disease is one of the most preventable causes of morbidity and mortality worldwide. For example, behavioural risk factors (including an unhealthy diet, physical inactivity, tobacco use and harmful use of alcohol) are responsible for about 80% of coronary heart disease and cerebrovascular disease (WHO, 2011). The World Cancer Research Fund (WCRF) estimates that about a third of the most common cancers are preventable through a nutritious diet, maintaining a healthy weight and regular physical activity (WCRF, 2014). The same lifestyle changes are known to reduce the risk of type 2 diabetes, with studies showing lifestyle intervention programs combining diet, exercise and modest weight loss can reduce the risk of developing diabetes by 58% (Knowler et al., 2002; Tuomilehto et al., 2001). The dietary recommendations for prevention of these common chronic diseases are similar and are consistent with the WHO 2004 *Global Strategy on Diet, Physical Activity and Health* which recommends increasing consumption of fruits, vegetables, legumes, whole grains and nuts, and reducing meat intake (WHO, 2004).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Meat is promoted as an important source of iron, zinc, vitamin B12, protein and omega-3s, and concerns are raised regarding nutritional inadequacies when it is excluded from the diet. Yet each of those nutrients are available from plant foods, with the exception of vitamin B12 (Zeuschner et al., 2013). It is interesting that despite a significant body of evidence showing health benefits of plant-based diets and negative health effects of diets high in red meat, that the focus continues to be on what is 'missing' when meat is removed from the diet. This is particularly the case in the Western world where chronic disease is a much bigger contributor to morbidity and mortality than nutrient deficiencies. Plant-based diets more closely match dietary recommendations for good health and prevention of chronic disease, and there is good evidence to show that plant-based diets containing little or no meat are associated with a reduced risk of chronic disease including cardiovascular disease, diabetes and some cancers. The relationship between red meat and chronic disease risk, the likely mechanisms underlying this relationship, and the reduction in chronic disease risk associated with diets low in red meat are explored in detail in this chapter.

A systematic search was conducted using the Medline database (www.pubmed.com; National Library of Medicine, Bethesda, MD) and Cochrane Reviews (www.cochrane.org/reviews; The Cochrane Collaboration) through to October 2014. Additional studies were obtained from reviewing the reference lists of other original studies and review papers. Due to the large number of studies, in most cases only the results of the most recent systematic reviews and meta-analyses have been included, along with other key papers where relevant and/or where a systematic review wasn't available.

RED MEAT INTAKE, HEALTH, AND CHRONIC DISEASE

Red Meat, Processed Meat, and Chronic Disease Risk

There is increasing evidence of an association between intake of red meat, and particularly processed meats, and chronic disease risk. This section explores the links between red meat intake and common chronic diseases including type 2 diabetes, cardiovascular disease, metabolic syndrome, cancer and obesity.

Type 2 Diabetes

At least 25 studies have been published assessing the relationship between meat intake and type 2 diabetes (T2D) risk, with the majority showing a positive association between red meat and/or processed meat intakes. The results of these studies have been incorporated into four meta-analyses (summarised in Table 1) and two combined analyses (of the Health Professionals Follow-Up Study (HPFS), Nurses' Health Study I (NHS-I) and Nurses' Health Study II (NHS-II)) published between 2009 and 2013.

Aune et al found that intakes of total meat, red meat and processed meat were each associated with an increased risk of T2D (Aune et al., 2009). The relative risks (RR) for the highest versus lowest intakes were 1.17 (95% confidence interval (CI) 0.92–1.48) for total meat, 1.21 (95% CI 1.07–1.38) for red meat and 1.41 (95% CI 1.25–1.60) for processed meat. Looking at dose response, they found that the relative risk per 120g/day red meat was 1.20 (95% CI 1.04–1.38) and per 50g/day processed meat was 1.57 (95% CI 1.28–1.93). A positive association between intakes of specific subgroups of meats (hamburger, bacon, hot dogs and other processed meats) and diabetes risk was also demonstrated. Micha et al, on the other hand, found no relationship between red meat intake and diabetes risk (RR 1.16; 95% CI 0.92-1.46), but

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 1. Summary of meta-analyses of studies on T2D risk and meat intake

Ref	Inc. Studies	Subjects	Cases	Meat Type	Measure 0f Intake	RR	95% CI	I^2	Heterogeneity p Value
Aune et al, 2009	n = 12 (all cohort studies)	445 323	6525	Total meat (n=5)	Highest vs lowest	1.17	0.92-1.48	86.9%	p<0.0001
		433 070	12 226	Red meat (n=10)	Highest vs lowest	1.21	1.07-1.38	58.5%	p=0.01
		380 606	9999	Processed meat (n=9)	Highest vs lowest	1.41	1.25-1.60	53.2%	p=0.03
				Total meat (n=4)	Per 120 g/day	1.26	0.84-1.88	90.6%	p<0.0001
				Red meat (n=9)	Per 120 g/day	1.20	1.04-1.38	68.3%	p=0.001
				Processed meat (n=8)	Per 50 g/day	1.57	1.28-1.93	74.0%	p<0.001
				Bacon (n=4)	Highest vs lowest	1.37	1.19-1.57	64.3%	p=0.04
				Hotdogs (n=5)	Highest vs lowest	1.30	1.20-1.42	0.0%	p=0.81
				Other processed meats (n=4)	Highest vs Lowest	1.25	1.09-1.44	63.9%	p=0.04
Micha et al, 2010	n = 7 (all cohort studies)	142 851	5923	Total meat (n=3)	Highest vs lowest	1.12	1.05-1.19	n/a	p=0.29
		298 982	7349	Red meat (n=5)	Highest vs lowest	1.16	0.92-1.46	n/a	p=0.25
		372 279	10 782	Processed meat (n=7)	Highest vs lowest	1.19	1.11-1.27	n/a	p<0.001
				Bacon (n=5)	Per serve (2 slce)/day	2.07	1.40-3.04	n/a	n/a
				Hotdogs (n=4)	Per serve (1 hotdog)/day	1.92	1.33-2.78	n/a	n/a
				Other processed meats (n=4)	Per serve (1 pce)/day	1.66	1.13-2.42	n/a	n/a
Pan et al, 2011	n = 9 (all cohort studies)	442 101	28 228	Total red meat	Highest vs lowest	1.34	1.26-1.43	n/a	n/a
				Unprocessed red meat	Highest vs lowest	1.25	1.17-1.33	n/a	n/a
				Processed meat	Highest vs lowest	1.32	1.24-1.40	n/a	n/a
				Unprocessed red meat (n=10)	Per 100g/day	1.19	1.04-1.37	93.3%	p=0.000
				Processed red meat (n=9)	Per 50g/day	1.51	1.25-1.83	94.3%	p=0.000
Feskens et al, 2013	n=24 (all cohort studies)	n/a	n/a	Total meat (n=14)	Per 100g/day	1.15	1.07-1.24	58%	n/a
				Red meat (n=14)	Per 100g/day	1.13	1.03-1.23	36%	n/a
				Unprocessed red meat (n=11)	Per 100g/day	1.15	0.99-1.33	94%	n/a
				Processed meat (n=21)	Per 50g/day	1.32	1.19-1.48	89%	n/a

n/a - Data not provided.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

did demonstrate a 19% higher risk of diabetes with higher intakes of processed meats (RR 1.19; 95% CI 1.11-1.27) (Micha et al., 2010).

In a combined analysis of findings from the HPFS, NHS-I and NHS-II, Pan et al. found a positive association between T2D risk and both unprocessed and processed red meat in each of the three cohorts (Pan et al., 2011). The pooled Hazard Ratios (HR) for a one serving per day increase in unprocessed, processed, and total red meat consumption were 1.12 (95% CI 1.08-1.16), 1.32 (95% CI 1.25-1.40), and 1.14 (95% CI 1.10-1.18), respectively. They estimated that replacing one serving (85g) per day of unprocessed red meat with one serving of nuts, low fat dairy or wholegrains could reduce the risk of T2D by 16-24% and replacing one serving of processed meat (28g bacon, 45g other processed meats) with the same foods could reduce the risk by 29-35%. The authors also conducted a meta-analysis, incorporating their findings into those of previous meta-analyses, which confirmed an increase in risk with both unprocessed (RR per 100g/day 1.19; 95% CI, 1.04-1.37) and processed meat (RR per 50g/day 1.51; 95% CI, 1.25-1.83) (Pan et al., 2011).

More recently, the same researchers explored the association between changes in red meat consumption and development of T2D in the same three study cohorts (Pan et al., 2013). They found that increasing red meat intake during a 4-year interval was associated with an increased risk of developing T2D in the subsequent 4 years. Compared with no change in intake, increasing meat red meat intake by more than 0.5 servings/day was associated with a 48% increased risk (pooled HR 1.48; 95% CI, 1.37-1.59). After adjustment for initial body mass index (BMI) and concurrent weight gain, the association was attenuated but remained significant (HR 1.30; 95% CI, 1.21-1.41) suggesting that the association may be partly mediated by weight. Interestingly, reducing red meat intake by at least 0.5 servings/day had no effect on risk in the first 4 years of follow-up, but reduced the risk by 14% (pooled HR 0.86; 95%CI 0.80-0.93) during a longer follow-up period (12-16 years).

The most recent meta-analysis, published in 2013, similarly found an association between higher intakes of total meat, unprocessed red meat and processed meat and T2D risk (Feskens et al., 2013). The pooled RR (95% CI) were 1.15 (1.07-1.24) per 100g of total meat, 1.13 (1.03–1.23) per 100g unprocessed red meat and 1.32 (1.19–1.48) per 50g processed meats. In addition to research associating red meat intake with risk of T2D, two further studies have shown that a higher intake of red meat prior to falling pregnant may increase the risk of developing gestational diabetes, a known risk factor for T2D which shares the same underlying pathophysiology (Zhang, Schulze, Solomon & Hu, 2006; Bao, Bowers, Tobias, Hu & Zhang, 2013; Bowers et al., 2011).

Cardiovascular Disease

Research has linked higher intakes of red meat and processed meat with an increased risk of cardiovascular disease including coronary heart disease and stroke (Bernstein et al., 2012; Bernstein et al., 2010; de Oliveira Otto et al., 2012; Kaluza, Akesson & Wolk, 2014; Kaluza et al., 2012; Kontogianni, Panagiotakos, Pitsavos, Chrysohoou & Stefanadis, 2008; Micha et al., 2012; Qi, van Dam, Rexrode & Hu, 2007). In 2010, Micha et al published a systematic review and meta-analysis of the effects of meat consumption and CHD risk (Micha et al., 2010). They identified four studies (three cohort and one case-control) evaluating the relationship between unprocessed red meat intake and six cohort studies evaluating processed meat intake and CHD. Their pooled analysis (see Table 2) found no significant association between unprocessed red meat intake and CHD risk (RR for each 100g/day serve: 1.00; 95% CI 0.81-1.23) but significantly increased risk with processed meat intake (RR for each 50g/day serve: 1.42, 95% CI 1.07-1.89).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 2. Summary of meta-analyses of Studies on CVD and red meat intake

Ref	Inc. Studies	Subjects	Cases	Type of Meat	Measure of Intake	RR	95% CI	I^2	Heterogeneity p Value
CHD									
Micha et al, 2010	n = 10 studies (8 cohort and 2 case-control)	635 558	22 562	Total meat (n=5)	Per 100g/day	1.27	1.07-1.89	n/a	P=0.002
		56 311	769	Red Meat (n=4)	Per 100g/day	1.00	0.81-1.23	n/a	P=0.36
		614 062	21 308	Processed meat (n=5	Per 50g/day	1.42	1.07-1.8	n/a	P=0.04
Stroke									
Micha et al, 2010	n = 3 (all cohort studies)	115 500	931	Total meat (n=2)	Per 100g/day	1.24	1.08- 1.43	n/a	n/a
		108 898	1700	Red meat (n=2)	Per 100g/day	1.17	0.40- 3.43	n/a	n/a
		108 898	1434	Processed meat (n=2)	Per 50g/day	1.14	0.94- 1.39	n/a	n/a
Kaluza et al, 2012	n = 6 (all cohort studies)	329 495	10 630	Total meat (n=5)	Per serve (100-120g)/day	1.11	1.06-1.16	0%	P=0.59
				Unprocessed red meat (n=6)	Per serve (100-120g)/day	1.11	1.03-1.20	37.8%	P=0.17
				Processed meat (n=5)	Per serve (50g)/day	1.13	1.03-1.24	0%	P=0.65

n/a - Data not provided.

Since this review, Bernstein et al. published findings from the NHS-I evaluating the association between unprocessed red meat and processed meat intake and CHD risk in women. They found a higher risk of both unprocessed red meat (RR per 100g/day: 1.19; 95% CI 1.07-1.32) and processed meat (RR per 50g/day: 1.20; 95% CI 1.03-1.40) intakes, after adjusting for potential confounding factors (Bernstein et al., 2010). Subsequently, Pan et al. combined the findings from the NHS-I, NHS-II and the HPFS to evaluate the relationship between red and processed meat intakes and CVD mortality (Pan et al., 2012). The pooled analysis demonstrated a higher risk of mortality from CVD for both unprocessed red meat (RR 1.18; 95% CI 1.13-1.23) and processed meat (RR 1.21; 95% CI 1.13-1.31) intakes.

Eight studies have explored the relationship between meat intake and stroke risk and the findings of these studies have been summarised in two meta-analyses published in 2010 and 2012 (see Table 2) (Kaluza et al., 2012; Micha et al., 2010). Micha et al. failed to find a relationship between red and processed meat intake and stroke, but identified only three studies, none of which evaluated the same type of meat (i.e. total versus unprocessed versus processed meat) and stroke subtype (Micha et al., 2010). More recently, Kaluza conducted a meta-analysis of six prospective cohort studies published between 2003 and 2012 (Kaluza et al., 2012). Of these, one study assessed mortality from stroke while five looked at stroke incidence, four of which assessed stroke subtypes in addition to total stroke. Combining the results of these studies they found that consumption of red meat, including both unprocessed red meat intake and processed meat was associated with an increased risk of total stroke and ischemic stroke but not haemorrhagic stroke. For ischemic stroke the RR (95% CI) were 1.13 (1.00-1.27) for unprocessed red meat, 1.15 (1.06-1.24) for processed meat and 1.12 (1.05-1.19) for total red meat. Bernstein et al., in their combined analysis of the NHS-I and HPFS, found that replacing one serving/day of red meat

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

with one serving of poultry, fish, nuts or dairy foods (both full fat and low fat) could reduce the risk of stroke by 10-27% (Bernstein et al., 2012).

Several other studies have explored the relationship between red meat intake and cardiovascular risks. In women with T2D, who are at higher risk of CVD, high intakes of both haem iron and red meat were positively associated with a risk of fatal CHD, coronary revascularisation and total CHD (Qi et al., 2007). In the multivariate analysis, the RR (95% CI) of CHD in the highest versus lowest quintiles of red meat intake were 2.05 (1.08-3.90) for fatal CHD, 1.91 (0.96-3.83) for coronary revascularisation and 1.36 (0.97-1.91) for total CHD. A similar increase in risk was seen for haem iron but not total iron intakes. In the Multi-Ethnic Study of Atherosclerosis (MESA), a positive association was found between CVD risk and intakes of haem iron from red meat (HR 1.65; 95%CI 1.10-2.47) and zinc from red meat (HR 1.51; 95% CI 1.02-2.24) (de Oliveira Otto et al., 2012). Finally, a recent study found that consumption of processed but not unprocessed red meat was associated with an increased risk of heart failure (HF) in men. Compared to those who consumed <25g/day, those who consumed more than 75g/day had a 28% higher incidence of HF (RR 1.28; 95% CI 1.10-1.48) and almost a 2.5-fold risk of HF mortality (RR 2.43; 95% CI 1.52-3.88) (Kaluza, Akesson, et al., 2014).

Metabolic Syndrome

While no meta-analyses have been conducted, several studies have demonstrated a link between red meat intake and risk of metabolic syndrome (MetS) (Azadbakht & Esmaillzadeh, 2009; de Oliveira Otto et al., 2012; Babio et al., 2012). Azadbakht et al., in a cross-sectional study of 482 Tehrani female teachers aged 40–60 years, found that compared to those in the lowest quintile of red meat intake, those with the highest intakes had a significantly greater odds of having MetS (odds ratio (OR) 2.15; 95% CI, 1.18-4.01) (Azadbakht & Esmaillzadeh, 2009). In the MESA, de Oliveira et al. followed 6814 adults, aged 45–84 years for up to 8-10 years. They found that intakes of haem iron and zinc from red meat, but not from other sources, were positively associated with risk of MetS (de Oliveira Otto et al., 2012). The HR were 1.25 (95% CI 0.99-1.56) for haem iron intake from red meat and 1.29 (95% CI 1.03-1.61) for zinc intake from red meat. In the PREDIMED (PREvención con DIeta MEDiterránea, or Prevention with Mediterranean Diet) study, subjects in the highest quartile of red meat intake were more likely to meet the criteria for MetS both at baseline (OR 2.3; 95% CI, 1.4-3.9) and after 1 year follow-up (OR 2.2; 95% CI, 1.3-3.7) compared with those with the lowest intakes, after adjusting for potential confounders (Babio et al., 2012). The risk of developing MetS during follow-up was also significantly higher in those with the highest versus lowest red meat intakes (OR 2.7; 95% CI, 1.1-6.8). Consistent with these findings, the CARDIA (Coronary Artery Risk Development in Young Adults) study demonstrated an association between higher intakes of animal but not vegetable protein and incidence of MetS (Pereira et al., 2002).

Cancer

Research has linked higher intakes of red meat and processed meat with an increased risk of some types of cancer but not others. Numerous meta-analyses have been published exploring the relationship between red and/or processed meat intakes and most types of cancer, the main findings of which are summarised in Table 3.

The most consistent findings are for colorectal cancer with six meta-analyses published between 2009 and 2014 showing an association between red and/or processed meat intake and colorectal cancer

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 3. Summary of meta-analyses evaluating the association between red and processed meat intakes and cancer risk

Ref	Inc Studies	Subjects	Cases	Type of Meat	Measure of Intake	RR	95% CI	I^2	Heterogeneity p Value
Bladder Cancer									
Wang, 2012	n = 20 (10 cohort, 11 case-control)	1 273740	8488	Total meat (n=11)	Highest vs lowest	1.04	0.80-1.27	62.3%	0.003
				Red meat (n=9)	Highest vs lowest	1.17	1.02-1.34	53.7%	n/a
				Processed meat (n=11)	Highest vs lowest	1.10	1.00-1.21	3.5%	n/a
Li, 2014	n = 15 (6 cohort, 9 case-control)	1 558 848	16 646	Red meat (n=14)	Highest vs lowest	1.15	0.97-1.36	73.5%	0.0000
				Processed meat (n=11)	Highest vs lowest	1.22	1.04-1.43	64.9%	0.002
Breast Cancer									
Alexander, 2010	n = 11 (includes 16 prospective cohorts (8 pooled) and 2 nested case-control	n/a	n/a	Red meat	Highest vs lowest	1.02 (FE)	0.98-1.07	n/a	0.001
				Processed meat	Highest vs lowest	1.07 (RE)	0.98-1.17	n/a	n/a
					Per 100 g	1.04 (FE)	1.00-1.07	n/a	0.0001
					Per 100g	1.12 (RE)	1.03-1.23	n/a	n/a
					Highest vs lowest	1.00 (FE)	0.98-1.01	n/a	0.005
					Highest vs lowest	1·08 (RE)	1.01-1.16	n/a	n/a
					Per 30g	1·03 (FE)	1.00-1.06	n/a	<0.0001
					Per 30g	1·06 (RE)	0.99-1.14	n/a	n/a
Colorectal Cancer									
Smolinksa, 2009	n = 22 (12 case-control, 10 cohort)	n/a	n/a	Red meat	More than 50g/day	1.21 (colon)	1.07-1.37	n/a	n/a
					More than 50g/day	1.30 (rectal)	0.90-1.89	n/a	n/a
					More than once/day	1.37 (colon)	1.09-1.71	n/a	n/a
					More than once/day	1.43 (rectal)	1.24-1.64	n/a	n/a
Chan, 2011	n = 24 (2 case-cohort, 3 nested case-control, 19 cohort)	n/a	n/a	Total Meat	Highest Vs Lowest	1.22	1.11-1.34	14%	0.31
				Red Meat	Highest vs Lowest	1.10	1.00-1.21	22%	0.22
				Processed Meat	Highest vs Lowest	1.17	1.09-1.25	6%	0.39
				Total Meat	Per 100g/ day	1.14	1.04-1.24	56%	0.01
				Red Meat	Per 100g/ day	1.17	1.05-1.31	0%	0.483
				Processed Meat	Per 50g/day	1.18	1.10-1.28	1%	0.333

continued on following page

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 3. Continued

Ref	Inc Studies	Subjects	Cases	Type of Meat	Measure of Intake	RR	95% CI	I^2	Heterogeneity p Value
Aune, 2013	n = 26 (19 case control, 7 prospective cohort)	234 451	21 493	Red Meat (n=16)	Per 100g/ day	1.27	1.16-1.40	5%	0.40
					Highest vs Lowest	1.22	1.11-1.34	46%	0.02
		41538	6098	Processed Meat (n=10)	Per 50g/day	1.29	1.10-1.53	27%	0.19
					Highest vs Lowest	1.29	1.15-1.45	18%	0.27
Johnson, 2013	n = 19 9 cohort, 10 case-control)	n/a	13 506	Red Meat (n=14)	5 servings/ week	1.13	1.09-1.16	n/a	n/a
			7963	Processed Meat (n=5)	5 servings/ week	1.09	0.93-1.25	n/a	n/a
Xu, 2013	n = 21 (16 case-control, 5 nested case-control/cohort)	92 057	2 793	Red Meat (n=21)	Highest vs Lowest	1.24	1.12-1.36	38.2%	0.036
				Processed Meat (n=7)	Per 100g/ day	1.36	1.17-1.58	69.2%	0.001
					Highest vs Lowest	1.17	1.08-1.26	0%	0.449
					Per 50g/day	1.28	1.03-1.60	69.1%	0.004
Pham, 2014	n = 14 (6 cohort, 8 case-control	566 597	5 957	Total Meat	Highest vs Lowest	1.06 (colorectal)	0.92-1.22	n/a	n/a
						1.17 (colon)	0.99-1.39	n/a	n/a
						0.90 (rectal)	0.71-1.14	n/a	n/a
				Red Meat	Highest vs Lowest	1.16 (colorectal)	1.00-1.34	28.1%	0.109
						1.21 (colon)	1.03-1.43	n/a	n/a
						1.05 (rectal)	0.81-1.37	n/a	n/a
				Processed Meat	Highest vs Lowest	1.17 (colorectal)	1.02-1.35	29.8%	0.086
						1.23 (colon)	1.03-1.47	n/a	n/a
						1.11 (rectal)	0.87-1.40	n/a	n/a
Endometrial Cancer									
Bandera, 2007	n = 19 (3 cohort, 16 case-control)	n/a	n/a	Total Meat	100g/day	1.26	1.03-1.54	79.6%	<0.01
					Highest vs Lowest	1.39	1.13-1.71	62.3%	<0.01
				Red Meat	100g/day	1.51	1.19-1.93	44%	0.97
					Highest vs Lowest	1.48	1.22-1.80	49.8%	0.06

continued on following page

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 3. Continued

Ref	Inc Studies	Subjects	Cases	Type of Meat	Measure of Intake	RR	95% CI	I^2	Heterogeneity p Value
Gastric Cancer									
Bonequi, 2013	n = 11 (case-control)	n/a	n/a	Total meat (n=5)	Highest vs Lowest	1.14	0.47-2.73	89.6%	<0.001
				Red Meat (n=5)	Highest vs Lowest	1.73	1.20-2.51	64.5%	0.024
				Processed Meat (n=6)	Highest vs Lowest	1.64	1.08-2.48	64.8%	0.015
Zhu, 2013	n = 42 (12 cohort, 30 case-control)	2 410 994	16 798	Red Meat (n=17)	Highest vs Lowest	1.45	1.22-1.73	76.4%	<0.001
				Processed meat (n=26)	Highest vs Lowest	1.45	1.26-1.65	61%	<0.001
				Beef (n=8)	Highest vs Lowest	1.28	1.04-1.57	0%	0.47
				Pork (n=5)	Highest vs Lowest	1.31	0.97-1.78	28.9%	0.162
				Bacon (n=7)	Highest vs Lowest	1.37	1.17-1.61	0%	0.695
				Ham (n=5)	Highest vs Lowest	1.44	1.00-2.06	77.9%	0.001
				Sausage (n=9)	Highest vs Lowest	1.33	1.16-1.52	59%	0.012
Stomach Cancer									
Song, 2014	n = 18 (4 cohort, 14 case-control)	1 228 327	n/a	Red Meat	Highest vs Lowest	1.37	1.18-1.59	67.6%	<0.001
					Per 100g/day	1.17	1.05-1.32	73.8%	<0.001
Kidney Cancer									
Alexander, 2009	n = 28 (12 case-control, 3 cohort, 1 pooled study of 13 cohort studies)	1 221 764	8 674	Red Meat (n=11)	Highest vs Lowest	1.12	0.98-1.29	n/a	0.015
				Processed Meat (n=10)	Highest vs Lowest	1.07	0.94-1.23	n/a	0.062
Renal Cell Cancer									
Lee, 2008	n = 13 (13 prospective cohort)	774 952	1 478	Red Meat	Q2 vs Q4	1.20 (woman)	0.93-1.55	n/a	n/a
					Q2 vs Q4	0.88 (men)	0.72-1.07	n/a	n/a
					Increase of 2 serves/wk	1.00	0.95-1.06	n/a	0.75
				Processed Meat	Increase of 2 serves/wk	1.01	0.99-1.02	n/a	0.96
Lung Cancer									
Yang, 2012	n = 34 (11 cohort, 23 case-control)	1 797 042	30 293	Total meat (n=2)	Highest vs Lowest	1.35	1.08-1.69	75.7%	<0.001
				Red Meat (n=18)	Highest vs Lowest	1.34	1.18-1.52	63.9%	<0.001
				Processed Meat (n=10)	Highest vs Lowest	1.06	0.90-1.25	79.5%	<0.001

continued on following page

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 3. Continued

Ref	Inc Studies	Subjects	Cases	Type of Meat	Measure of Intake	RR	95% CI	I^2	Heterogeneity p Value
Xue, 2014	n = 33 (27 case control and 6 cohort)	n/a	n/a	Red Meat (n=28)	Highest vs Lowest	1.44	1.29-1.61	77.0%	0.000
				Red Meat (n=11)	Per 120g serve/day	1.35	1.25-1.46	n/a	n/a
				Processed Meat (n=23)	Highest vs Lowest	1.23	1.10-1.37	75.7%	0.000
				Processed Meat (n=11)	Per 50g serve/day	1.20	1.11-1.29	n/a	n/a
Oral Cavity and Oropharynx Cancer									
Xu, 2014	n = 13 (1 cohort, 12 case-control)	501 730	4 104	Total Meat (n=9)	Highest vs Lowest	1.14	0.78-1.68	82.9%	<0.001
				Red Meat (n=3)	Highest vs Lowest	1.05	0.66-1.66	49.4%	0.12
				Processed Meat (n=9)	Highest vs Lowest	1.91	1.19-3.06	85.9%	<0.001
Oesophageal Cancer									
Salehi, 2013	n = 35 (4 cohort, 31 case-control)	1 591 128	10 818	Total meat (n=21)	Highest to Lowest	0.99	0.85-1.15	49.7%	0.005
				Red Meat (n=14)	Highest to Lowest	1.40	1.09-1.81	n/a	0.001
				Processed Meat (n=17)	Highest to Lowest	1.41	1.13-1.76	62.0%	0.0001
				Barbequed Meat (n=3)	Highest to Lowest	1.68	0.88-3.21	n/a	0.04
				Total meat (n=9)	Per 100g/ day	1.01	0.99-1.01	33.8%	0.02
				Processed Meat (n=8)	Per 50g/day	1.57	1.22-2.01	n/a	n/a
Huang, 2013	n = 10 (3 cohort, 7 case-control)	1 146 096	2 037	Red Meat (n=9)	Highest vs Lowest	1.31	1.05-1.64	18.9%	0.274
				Processed Meat (n=9)	Highest vs Lowest	1.41	1.09-1.83	39.4%	0.105
				Red meat (n=7)	Per 100g/ day	1.45	1.09-1.93	61.8%	0.015
				Processed Meat (n=7)	Per 50g/day	1.37	1.03-1.81	71.0%	0.002
Choi, 2013	n = 27 (4 cohort, 23 case-control)	1 176 331	7 489	Red Meat (n=22)	Highest vs Lowest	1.38	1.17-1.64	67.1%	<0.01
				Processed Meat (n=18)	Highest vs Lowest	1.32	1.08-1.62	58.4%	<0.01
				Red Meat (n=3)	Per 100g/ day	1.05	0.91-1.21	0.2%	0.42
				Processed Meat (n=3)	Per 100g/ day	1.37	0.88-2.13	33.5%	0.17

continued on following page

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 3. Continued

Ref	Inc Studies	Subjects	Cases	Type of Meat	Measure of Intake	RR	95% CI	I^2	Heterogeneity p Value
Zhu, 2014	n = 35 (7 cohort, 28 case-control)	1 919 078	13 313	Total Meat (n=24)	Highest vs Lowest	1.19	0.98-1.46	73.3%	<0.001
				Red Meat (n=15)	Highest vs Lowest	1.55	1.22-1.96	63.6%	<0.001
				Processed Meat (n=15)	Highest vs Lowest	1.33	1.04-1.69	61.5%	<0.001
Ovarian Cancer									
Kohladooz, 2010	n = 17 (6 cohort, 11 case-control)	730 836	6 720	Total Meat (n=7)	Highest vs Lowest	1.16	0.96-1.40	n/a	0.04
				Red Meat (n=9)	Highest vs Lowest	1.16	1.02-1.32	n/a	0.07
				Processed Meat (n=5)	Highest vs Lowest	1.19	1.07-1.34	n/a	0.88
Wallin, 2011	n = 8 (all cohorts)	754 836	2 349	Total Meat (n=8)	Per 100g/ week	1.02	1.00-1.05	0.0%	0.96
				Red Meat (n=8)	Per 100g/ week	1.02	0.99-1.04	0.0%	0.97
				Processed Meat (n=5)	Per 100g/ week	1.05	0.98-1.14	0.0%	0.65
				Red Meat (n=8)	Per 4 x 100g serves/wk	1.07	0.97-1.19		
				Processed Meat (n=5)	Per 4 x 30g serves/wk	1.07	0.97-1.17		
Pancreatic Cancer									
Paluszkiewicz, 2012	n = 11 (6 cohort, 5 case-control)	947 460	3 511	Red Meat	Highest to Lowest	1.27	1.07-1.50	73.1%	<0.0001
Larsson, 2012	n = 11 (all cohort)	2 307 787	6 643	Red Meat (n=11)	Per 120g/ day	1.13	0.93-1.39	69.8%	<0.001
				Processed Meat (n=7)	Per 50g/day	1.19	1.04-1.36	0.0%	0.46
Prostate Cancer									
Alexander, 2010	n = 17 (all cohort)	n/a	n/a	Red Meat (n=15)	Highest vs Lowest	1.00	0.96-1.05	n/a	0.264
				Processed Meat (n=11)	Highest vs Lowest	1.05	0.99-1.12	n/a	0.088
				Red Meat (n=9)	Per 100g	1.00	0.95-1.05	n/a	0.007
				Processed Meat (n=10_	Per 30g	1.02	1.00-1.04	n/a	0.274

n/a - Data not provided.
FE= Fixed effects model; RE = Random effects model.

(CRC) risk (Aune et al., 2013; Chan et al., 2011; Johnson et al., 2013; Pham et al., 2014; Smolinska & Paluszkiewicz, 2010; Xu et al., 2013). In the largest and most recent of these, Aune et al. combined the results of 26 studies, (19 case-control and 7 prospective cohort studies) as part of the Continuous Update Project of the World Cancer Research Fund (Aune et al., 2013). They found an increased risk of

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

CRC associated with both red (RR per 100g 1.27; 95% CI 1.16-1.40) and processed meat (RR per 50g/day 1.29; 95% CI 1.10-1.53) intakes. These findings are very similar to those of a second meta-analysis published around the same time, incorporating 21 studies (16 case-control and five cohort studies). The RR of CRC in this analysis was 1.36 (95% CI 1.17-1.58) for every 100g increase in red meat intake and 1.28 (95% CI 1.03-1.60) per 50 g/day increase in intake of processed meat (Xu et al., 2013). An earlier meta-analysis found a higher risk of CRC with an increasing frequency of red meat consumption than with the total amount consumed (Smolinska & Paluszkiewicz, 2010).

An association between red and/or processed meat intakes has also been found for several other gastrointestinal cancers including oesophageal (Hua, Stoohs & Facchini, 2001; Salehi, Moradi-Lakeh, Salehi, Nojomi & Kolahdooz, 2013; Zhu et al., 2014), gastric (Bonequi, Meneses-Gonzalez, Correa, Rabkin & Camargo, 2013; Song et al., 2014; Zhu et al., 2013), and oral and oropharynx (Xu, Yang, Wu, Li & Bai, 2014). The most recent of four meta-analysis for oesophageal cancer combined the findings of 35 studies (7 cohort and 28 case-control), showing an increased risk for the highest versus lowest categories of intake for total meat (RR 1.19; 95% CI 0.98-1.46), red meat (RR 1.55; 95% CI 1.22-1.96) and processed meat (RR 1.33; 95% CI 1.04-1.69) (Zhu et al., 2014). For gastric cancer, the most recent meta-analysis found a significant increase in risk (RR 1.37; 95% CI 1.18-1.59) for the highest versus lowest categories of red meat intake but didn't look separately at processed meats (Song et al., 2014). A previous meta-analysis combining the findings of 12 cohort and 30 case-control studies showed a positive association between gastric cancer risk and intakes of both red and processed meats as well as specific meat subgroups including beef, ham, bacon and sausages (Zhu et al., 2013). Xu et al. found a significant association between intakes of processed meats and oral cavity and oropharynx cancers (RR 1.91, 95% CI 1.19-3.06) but no association with total meat or red meat intakes (Xu et al., 2014).

In addition to gastrointestinal cancers, there is evidence of an association between red meat intakes and endometrial (Bandera, Kushi, Moore, Gifkins & McCullough, 2007), lung (Xue et al., 2014; Yang et al., 2012) and pancreatic (Larsson & Wolk, 2012; Paluszkiewicz, Smolinska, Debinska & Turski, 2012) cancers and between processed meat intakes and bladder, lung (Xue et al., 2014), ovarian, and pancreatic (Larsson & Wolk, 2012) cancers. While an earlier meta-analysis of case-control studies found a relationship between total, red and processed meat intakes and renal cancer, two more recent meta-analyses have failed to show an association (Alexander & Cushing, 2009; Lee et al., 2008). While some individual studies have shown associations with breast and prostate cancer, the most recent meta-analyses don't support a relationship between meat intake and these cancers (Alexander, Morimoto, Mink & Cushing, 2010; Alexander, Mink, Cushing & Sceurman, 2010). However there is some suggestion of a stronger relationship between meat intake and breast cancer in postmenopausal women(Alexander, Morimoto, et al., 2010).

Obesity

Red meat intake has been positively associated with obesity and central obesity (Wang & Beydoun, 2009) and several studies have found an association between a higher intake of red meat and an increased risk of weight gain (Romaguera et al., 2010; Schulz et al., 2002; Vergnaud et al., 2010; Mozaffarian, Hao, Rimm, Willett & Hu, 2011). In the European Prospective Investigation into Cancer and Nutrition–Physical Activity, Nutrition, Alcohol Consumption, Cessation of Smoking, Eating Out of Home, and Obesity (EPIC-PANACEA) study, those with high adherence to a Mediterranean dietary pattern gained less weight and were less likely to develop overweight and obesity, an effect that was largely explained by the low

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

meat content of a Mediterranean style diet (Romaguera et al., 2010). When the researchers assessed the relative contribution of each component of the MD score, the association between the MD score and weight change was no longer significant after excluding meat and meat products. These findings are consistent with research showing that vegetarians tend to have a lower BMI and experience less weight gain than meat-eaters (Rosell et al., 2006; Tonstad, Butler, Yan & Fraser, 2009).

Red Meat Intake and Mortality

Considering the links with chronic disease risk, it is perhaps not surprising that higher intakes of red meat have also been associated with a higher risk of mortality. A large study investigating the association of a wide range of meat intakes with chronic disease mortality found that both red and processed meat intakes were associated with modest increases in total mortality, cancer mortality, and cardiovascular disease mortality (Sinha et al., 2009). The EPIC study showed a moderate positive association between processed meat consumption and mortality, in particular due to cardiovascular diseases, but also to cancer (Rohrmann et al., 2013). A combined analysis of findings from the HPFS and NHS demonstrated an association between red meat consumption and an increased risk of total, CVD, and cancer mortality and that substituting red meat with other protein sources (including fish, poultry, nuts, legumes, low-fat dairy, and whole grains) is associated with a lower mortality risk (Pan et al., 2012).

A recent meta-analysis found that high consumption of red meat, especially processed meat, may increase all-cause mortality (Larsson & Orsini, 2014) while another found that processed meat consumption may increase the risk of mortality from any cause and CVD, while red meat consumption was positively but weakly associated with CVD mortality (Abete et al., 2014). A low meat intake, on the other hand, has been associated with an increased life expectancy, particularly with increasing duration of adherence (Singh, Sabate & Fraser, 2003).

Mechanisms to Explain the Association between Red Meat and Chronic Disease Risk

There is growing evidence to suggest that the adverse effects of meat on health are related not just to saturated fat and cholesterol (as is commonly assumed), but to other substances present in meat. These include nutrients in red meat (such as protein, amino acids, haem iron, L-carnitine and choline), substances produced during the processing and cooking of red meat (including heterocyclic amines, polycyclic aromatic hydrocarbons, nitrates, nitrites, nitrosamines and advanced glycation end products) and other components found almost exclusively in red meat such as NeuGc5 (Feskens et al., 2013; Micha, Michas, Lajous & Mozaffarian, 2013; de Oliveira Otto et al., 2012; Cross et al., 2010). Notably, these different compounds may act synergistically in increasing risk of disease (Kim, Coelho & Blachier, 2013). This section will explore the recent evidence on each of these components to understand how red meat can increase our risk of chronic diseases such as cardiovascular disease, type 2 diabetes mellitus and some cancers.

Saturated Fat

Red meat is a significant source of saturated fat (SF) in the diet and is considered one of the main reasons for the link between red meat intake and disease risk, particularly cardiovascular risk. Saturated fat

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

contributes an average of 11 percent of total energy to the American diet, with meats such as sausages, franks, bacon and ribs contributing 5%, burgers 4.4% and beef/beef mixed dishes 4.1% (NCI, 2013). The American Heart Association recommends aiming for just 5% to 6% of total energy intake from saturated fat (Lichtenstein et al., 2006) and the US Dietary Guidelines recommend consuming less than 10% of calories from saturated fat (USDA & USDHHS, 2010).

The amount and quality of dietary fat and its effects on chronic disease have been studied intensively during the past decades. While previously low-fat diets were recommended without much attention to the type of fat, current guidelines place more emphasis on the quality of fat consumed (Schwab et al., 2014). Despite the recent controversy regarding the role of SF in heart disease, the bulk of evidence continues to recommend replacing SF with PUFA (poly-unsaturated fat) to reduce the risk of heart disease (Kromhout, Geleijnse, Menotti & Jacobs, 2011; Astrup et al., 2011; Willett & Stampfer, 2013; Farvid et al., 2014; Schwab et al., 2014). Considered one of the best systematic reviews examining SF and heart disease, Jakobsen et al. found that substitution of 5% of energy from SF with 5% from PUFA was associated with a significant reduction in CHD events and deaths (Jakobsen et al., 2009). These findings are supported by more recent meta-analyses. Mozaffarian et al. found that a 5% increase in PUFA, in place of SF, reduces CHD events by 10% (Mozaffarian, Micha & Wallace, 2010). Similarly, Farvid et al. showed that substituting 5% of SF with 5% of PUFA lowered risk of CHD events by 9% and risk of death from CHD by 13%(Farvid et al., 2014). Jakobsen et al. also reported that if SF is replaced with low glycemic index (GI) CHO, the risk of CHD is also decreased whereas replacing SF with high GI CHO is associated with higher risk, showing the importance of the overall quality of both fats and CHO in the diet (Jakobsen et al., 2010).

A number of studies have demonstrated an association between high fat diets and risk of T2D (Colditz et al., 1992; Marshall, Hamman & Baxter, 1991; Marshall, Hoag, Shetterly & Hamman, 1994; Tsunehara, Leonetti & Fujimoto, 1990; Lindstrom et al., 2006). In the Finnish Diabetes Study, subjects with the highest fat intakes were more than twice as likely to develop T2D compared to those with the lowest fat intakes, while a higher intake of SF increased the risk of developing diabetes 1.7 times (Lindstrom et al., 2006). The mechanisms linking SF intake and diabetes risk appear to relate to insulin sensitivity, with several studies showing an association between SF intake and reduced insulin sensitivity (Feskens & Kromhout, 1990; Lovejoy & DiGirolamo, 1992; Marshall, Bessesen & Hamman, 1997; Mayer, Newman, Quesenberry & Selby, 1993; Mayer-Davis et al., 1997; Parker et al., 1993). A higher SF intake increases the proportion of SF in muscle cell membranes, which possibly influences insulin action by altering insulin receptor binding or affinity and transport and cell signalling (Marsh & Brand-Miller, 2011; Parillo & Riccardi, 2004; Vessby, 2000).

Dietary Cholesterol

Until recently, dietary cholesterol was thought to play a role in heart disease, but a direct association between dietary cholesterol and risk of CHD has been observed only in a few cohorts (Djousse & Gaziano, 2009). Instead, dietary patterns high in total fat, SF and cholesterol, with a low intake of fibre and PUFA are more consistently correlated with risk of CHD. Whether this is due to the high consumption of SF or cholesterol or both, or perhaps an insufficient supply of one or more protective factors such as fibre or PUFA, has been difficult to determine (Kratz, 2005). A meta-analysis of randomized clinical trials showed only a minimal increase (0.02 units per 100mg) in the total/high-density lipoprotein cholesterol ratio, which suggests the association between dietary cholesterol and CHD risk is, if anything, minor in

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

nature (Kratz, 2005; Weggemans, Zock & Katan, 2001). However, lowering dietary cholesterol content might reduce the risk of CHD considerably in a subgroup of individuals who are highly responsive to changes in cholesterol intake (Kratz, 2005). Furthermore, one study found that the addition of a moderate amount of dietary cholesterol to a reduced fat diet rich in PUFA or SF increased in vitro susceptibility of low density lipoprotein (LDL) cholesterol to oxidation (Schwab et al., 2000).

Trans-Fats

Trans-fats (TF) increase the risk of CHD even more strongly than SF by raising LDL-cholesterol and total: HDL cholesterol and promoting inflammation (Mozaffarian et al., 2004; Baer, Judd, Clevidence & Tracy, 2004; Ascherio, Katan, Zock, Stampfer & Willett, 1999; Mensink, Zock, Katan & Hornstra, 1992; Kromhout et al., 2011). They have also been shown to reduce the particle size of LDL-cholesterol, increase blood levels of Lipoprotetin-a Lp(a) and adversely affect endothelial function (Mozaffarian, Katan, Ascherio, Stampfer & Willett, 2006). Furthermore, TF worsen insulin sensitivity, particularly those with pre-existing insulin resistance, visceral adiposity or lower physical activity (Mozaffarian, Aro & Willett, 2009)

Most TF in the western diet occur as the result of processing, formed during the partial hydrogenation of vegetable oils (converting them to solid or semi-solid fats for cooking or baking). The fat in red meat (and butter) also has small amounts (typically 4-5%) of trans-fatty acids due to bio-hydrogenation in the rumen of cattle and other ruminants (Willett, 2012). As the TF of processed foods have been reduced by manufacturers, TF from ruminant sources has become of greater interest, particularly in relation to the risk of CHD, however four prospective studies did not identify a significant positive association (Willett et al., 1993; Pietinen et al., 1997; Oomen et al., 2001; Jakobsen, Overvad, Dyerberg & Heitmann, 2008). The evidence from observational studies also suggests that higher CHD risk is related to consumption of industrially produced TF rather than ruminant TF. A WHO scientific update (Mozaffarian et al., 2009) on TF stated that the effects of ruminant and industrial TF are similar when consumed in similar quantities, but in the amounts consumed in actual diets, ruminant TF do not appear to be major contributors to CHD risk. However TF from ruminant sources may be of greater importance when consumed on top of TF from industrial sources (Mozaffarian et al., 2009).

Animal Protein

There is some evidence that animal protein itself may increase disease risk, with several studies showing an association between animal protein, but not plant protein, and risk of chronic disease. In healthy men (those without common cardiovascular risk factors) a higher intake of total and animal protein was associated with an increased risk of ischaemic heart disease (IHD), while there was no association with plant protein (Preis, Stampfer, Spiegelman, Willett & Rimm, 2010). Similarly, in an elderly Greek population, higher intakes of total protein and protein from meat and meat products were associated with a higher prevalence of diabetes while there was no association with protein from vegetables or cereals (Pounis et al., 2010) In the EPIC study, subjects with the highest intakes of animal protein had more than twice the risk of developing T2D compared to those with the lowest intakes while there was a 16% non-significant reduction in risk with vegetable protein (Sluijs et al., 2010). The researchers estimated that replacing 5% of energy from fat or carbohydrate with 5% of energy from animal protein would increase diabetes risk by about 20%. Vegetable or plant protein has also been inversely related to blood

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

pressure (Elliott et al., 2006; Altorf-van der Kuil et al., 2012). The reasons for the association between animal protein and disease risk are unclear but emerging data suggests that a higher intake of branched chain amino acids (leucine, isoleucine, valine) may increase insulin resistance, a risk factor underlying many chronic diseases (Beasley & Wylie-Rosett, 2013).

Despite the interest in high protein diets (typically from animal protein) for weight loss, there is some evidence that animal protein (excluding dairy and fish) is positively associated with long-term weight gain (Halkjaer et al., 2010; Clifton, 2011). Furthermore, long-term adherence to high protein diets, without discrimination towards protein source, may have adverse health consequences, as described above. Consistent with these findings, Fung et al (2010), in a combined analysis of the NHS and HPFS, found a higher animal-based low-carbohydrate intake was associated with a 23% higher all-cause mortality, 14% higher cardiovascular mortality and 28% higher cancer mortality comparing extreme deciles of intake, whereas a higher vegetable-based low-carbohydrate intake was associated with a 20% lower all-cause mortality and 23% lower cardiovascular mortality (Fung, van Dam, et al., 2010). A recent systematic review similarly found an increased risk of mortality with low carbohydrate diets (which by default are typically high in animal protein) however they didn't look specifically at the relationship with protein intake (Noto, Goto, Tsujimoto & Noda, 2013).

With respect to CRC risk, animal protein (as well as substances typically found in animal protein including haem iron, *N*-nitroso compounds, heterocyclic amines) can affect the large intestine mucosa with genotoxicity and metabolic disturbances (Kim et al., 2013). Increased bacterial fermentation (putrefaction) of undigested protein and production of bacterial metabolites derived from amino acids may affect colon epithelial homeostasis and renewal. Most colonic cancers are found in the distal colon and rectum where protein fermentation occurs (Kim et al., 2013).

Haem Iron

Red meat is typically recommended as an important source of haem iron, a cofactor for several enzymes and a major component of oxygen transporters in the body. Haem iron (40% of the total iron in meat) is easily absorbed in our body regardless of our body stores, whereas non-haem iron (the type of iron present in plant foods) is absorbed according to need (Saunders, Craig, Baines & Posen, 2012). A strong pro-oxidant and catalyst in many cellular reactions, haem iron causes tissue damage, and increases oxidative stress (Rajpathak, Ma, Manson, Willett & Hu, 2006; Kunutsor, Apekey, Walley & Kain, 2013; Broedbaek et al., 2011). As we have limited means of excreting excess iron, a high intake of haem iron and elevated serum ferritin concentrations (high iron stores) may be a risk factor for chronic disease, including heart disease, diabetes and some types of cancers.

Both high haem iron intake and elevated serum ferritin contribute to the development of atherosclerosis by catalysing the production of free radicals and promoting low density lipoprotein (LDL) oxidation (Hunnicutt, He & Xun, 2014). A recent meta-analysis found that participants with higher haem iron intakes had a 31% increased risk of CHD compared to those with lower intakes. An increase in haem iron intake of 1mg/day was associated with a 27% increase in risk of CHD (Yang et al., 2014). A study of postmenopausal women with T2D found that those with the highest intake of haem iron had a 50% increased risk of total CHD compared to those with the lowest intake (Qi et al., 2007). A high haem iron intake has also been associated with an increased risk of stroke (Kaluza, Wolk & Larsson, 2013) and fatal acute myocardial infarct (AMI) (Kaluza, Larsson, Hakansson & Wolk, 2014).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

There is consistent evidence showing a positive association between dietary haem iron intake from red meat and the risk of T2D (Luan de et al., 2008; Jiang et al., 2004; Rajpathak et al., 2006; Lee, Folsom & Jacobs, 2004; Fernandez-Cao et al., 2013), although one study found haem iron intake to be associated with CVD and metabolic syndrome but not T2D (de Oliveira Otto et al., 2012). Two recent meta-analyses have shown both increased ferritin levels and haem-iron intake are significantly associated with higher risk of T2D (Kunutsor et al., 2013; Zhao et al., 2012) where individuals in the top quintile of ferritin levels and dietary haem iron status have about a 70% and 30% higher risk, respectively, of T2D events compared to the bottom quintile (Kunutsor et al., 2013). Higher intakes of dietary haem iron during pre-pregnancy and early pregnancy have also been associated with an increased risk of gestational diabetes (Bowers et al., 2011; Qiu et al., 2011). The damaging effects of elevated iron stores on pancreatic β-cells and hepatic function impairs glucose and insulin metabolism leading to the development of T2D (Opara, 2004; Rajpathak et al., 2009; Zhao et al., 2012; Kunutsor et al., 2013; Bowers et al., 2011). Catalytic iron damages DNA and the integrity of cell membranes (iron-induced oxidative stress), and also interferes with glucose uptake into skeletal muscles and adipocytes, inducing insulin resistance (Zhao et al., 2012; Andrews, 1999; Lenzen, 2008).

Hereditary hemochromatosis is known to contribute to diabetes due to the excessive accumulation of iron in tissues, but reducing body iron levels using phlebotomy improves insulin sensitivity (Rajpathak et al., 2009). One study found vegetarians, who had lower iron stores (ferritin levels about half that of meat-eaters), were more insulin sensitive than meat-eaters. When six of the male meat-eaters had their iron stores lowered by phlebotomy to levels similar to that seen in vegetarians, they experienced a 40% increase in insulin sensitivity (Hua et al., 2001). Several studies have suggested that elevated levels of ferritin may help identify individuals at high risk of T2D (Kunutsor et al., 2013; Bao, Rong, Rong & Liu, 2012). The National Institutes of Health (NIH) currently defines elevated ferritin levels as ≥300ng/mL for men and ≥150ng/mL for women (NIH, 2014).

A recent meta-analysis found that higher intakes of haem iron have a positive association with cancer risk, particularly colorectal, colon, breast and lung cancers (Fonseca-Nunes, Jakszyn & Agudo, 2014). Dietary haem iron influences cancer progression (Hooda, Shah & Zhang, 2014) mainly due to its pro-oxidant activity (Fonseca-Nunes et al., 2014). Recent studies of lung cancer cells showed they required increased intracellular haem biosynthesis and uptake to meet the increased demands for oxygen-utilizing hemoproteins (Hooda et al., 2014). In the case of CRC, haem iron has a catalytic and toxic effect on both the endogenous formation of carcinogenic N-nitroso compounds (NOC) and the formation of lipid oxidation end products. Most of these NOC can yield alkylating agents during metabolism that cause DNA damage (Bastide, Pierre & Corpet, 2011; Kim et al., 2013). A recent observational study found that haem iron was associated with an increased risk of colorectal tumours and overexpression of P53, suggesting alkylating rather than oxidative DNA-damaging mechanisms are involved in haem-induced colorectal carcinogenesis (Gilsing et al., 2013).

Heterocyclic Amines and Polycyclic Aromatic Hydrocarbons

Heterocyclic amines (HCAs) and polycyclic aromatic hydrocarbons (PAHs) are known mutagens and probable human carcinogens that are formed when cooking meat at high temperatures, such as pan frying or cooking over an open flame (Sinha et al., 2005; Zheng & Lee, 2009). Both HCAs and PAHs occur in grilled meats and poultry, as well as in gravies made from dripping of well-done roasts and fish.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

More than 20 HCAs have been identified in red and processed meats and poultry (Kim et al., 2013). They are produced during cooking from the reaction of creatine or creatinine, amino acids and sugars, with greater amounts forming at higher temperatures and longer periods of time. How much is formed also depends on the type of meat and cooking methods. The most mass-abundant HCAs detected in cooked meats are PhIP (2-amino-1-methyl-6-phenylimidazo[4,5-*b*]pyridine) and MeIQx (2-amino-3,8-dimethylimidazo[4,5-*f*]quinoxaline). Animal studies have shown HCAs increase cancer risk in multiple sites, including mammary glands, lung, colon, stomach and prostate. More recently HCA-DNA adducts have been seen in a variety of human tissues, including breast, colorectum and prostate (Zheng & Lee, 2009).

There are 16 PAHs that have potentially genotoxic and carcinogenic effects on humans. They are formed in meats during grilling/roasting/barbequing, when fat and juices from the meat drip onto the fire causing flames. These flames contain PAHs that adhere to the surface of the meat. Other food preparation processes during which PAHs can form are drying, smoking or adding smoke flavour (Cross & Sinha, 2004). The main mechanism involves their effects on enzymes, especially on the cytochrome P450 (CYP) family of enzymes. Benzo(a)pyrene (B(a)P) is the most carcinogenic of the 16 PAHs and is generally used as a marker of the effect of PAHs in foods (Chen et al., 2012). The adducts of PAH-DNA tend to concentrate in the epithelium of the oesophagus, cervix and vulva, the prostate and the placenta (Pratt et al., 2011).

Zheng et al. recently reviewed 21 epidemiological studies published since 1996 specifically evaluating the association of meat carcinogen exposure (by high temperature cooking and meat doneness level) and cancer risk. Most of these studies found that high intakes of well-done meat and high exposure to meat carcinogens, particularly HCAs may increase the risk of human cancers including CRC, breast cancer, prostate cancer, pancreatic cancer, oesophageal and gastric cancers (Zheng & Lee, 2009). A more recent study has highlighted the combined genotoxic potential of B(a)P and PhIP on intestinal cell lines which reflect early steps of colorectal carcinogenesis (Jamin et al., 2013). Cross et al. found those with the highest intake of the heterocyclic amine DiMeIQx were at increased risk for gastric cardia cancer, whereas there was only a suggestive increased risk of oesophageal cancer from other HCAs (MeIQX, PhIP) or haem iron (Cross et al., 2011). Intakes of B(a)P and PhIP in red meat are also associated with an increased risk of renal cell carcinoma (Daniel et al., 2012).

N-Nitroso Compounds

Humans are exposed to N-nitroso compounds (NOCs), also known as nitrosamines, mainly through diet (processed meats), cigarette smoking, occupational exposure and endogenous formation in the human gut, the latter being influenced by the haem iron, nitrate and nitrite content of consumed meats. As previously mentioned, haem iron is a catalyst for the formation of endogenous NOC, causing DNA damage (Bastide et al., 2011; Kim et al., 2013). Although NOC are found in processed meats, they are also formed in the gut endogenously due to the nitrate and nitrite content of processed meat (added for preservation and enhancement of colour and flavour). Hence, red and processed meats are the most significant dietary components linked to the formation of NOC in the gut.

Although nitrate and nitrite are not carcinogenic themselves, under conditions that result in endogenous nitrosation, ingested nitrate and nitrite may possibly be carcinogenic to humans, increasing risk of CRC (Habermeyer et al., 2014; Dellavalle et al., 2014). Studies have shown that high protein diets (either low or moderate in CHO) result in increased intake of N-nitroso compounds with a significant

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

decrease in faecal cancer-protective metabolites and increase in concentrations of hazardous metabolites, with potential for an increased risk of colonic disease in the long-term (Russell et al., 2011). In addition, increased endogenous NOC formation combined with prolonged transit times in the gut may explain the epidemiological associations between high meat/low fibre diets and CRC risk (Hughes, Cross, Pollock & Bingham, 2001). Consumption of a meat-based high protein weight loss diet showed increased faecal NOC, with red meat intake positively correlating with the faecal log NOC concentrations (Holtrop, Johnstone, Fyfe & Gratz, 2012). Bladder cancer risk has also been linked with total dietary nitrite and nitrate plus nitrite from processed meats (Ferrucci et al., 2010) and a Dutch study found that N-nitroso compounds, haem iron intake and nitrite intake increased the risk of oesophageal squamous cell carcinoma in men, but not women (Keszei, Goldbohm, Schouten, Jakszyn & van den Brandt, 2013).

Nitrosamines have been found to be toxic to pancreatic beta cells, increasing the risk of type 1 and type 2 diabetes in animal studies (Helgason, Ewen, Ross & Stowers, 1982; Portha, Giroix, Cros & Picon, 1980) and type 1 diabetes in some epidemiological studies (Dahlquist, Blom, Persson, Sandstrom & Wall, 1990; Sipetic et al., 2005). This is consistent with evidence showing a higher risk of type 2 diabetes associated with processed (red) meat versus unprocessed meats (Feskens et al., 2013; Micha et al., 2012).

L-Carnitine, Choline, and TMAO

Recent research has discovered a link between L-carnitine and choline ingestion, gut microbiota metabolism and CVD risk which suggests gut microbiota play a role in development of atherosclerosis (Koeth et al., 2013). Choline is found in a range of foods but particularly in animal proteins including eggs, red meat, poultry and seafood. The richest source of L-carnitine is red meat. These trimethylamine containing compounds are metabolised by the gut microbiota to trimethylamine (TMA), then converted in the liver to trimethylamine-N-oxide (TMAO). Circulating TMAO levels are strongly associated with atherosclerosis (possibly via suppression of reverse cholesterol transport, cholesterol accumulation and foam cell formation) and may therefore be a potential predictor of CVD risk (Bennett et al., 2013; Wong, 2014; Brugere et al., 2014; Koeth et al., 2013; Wang et al., 2011).

A diet rich in red meat (a significant source of L-carnitine and choline) results in an increased production of TMAO (Koeth et al., 2013). While intakes of choline and L-carnitine are lower in vegetarian diets, a recent study also found that vegans and vegetarian produced less TMAO compared to omnivores following ingestion of L-carnitine due to differences in gut bacteria (Koeth et al., 2013). The same authors found that plasma L-carnitine levels predicted increased risks of both CVD and major cardiac events (myocardial infact, stroke or death) but only among subjects with concurrently high TMAO levels. The lower dietary intake of L-carnitine and total choline by vegans and vegetarians and their reduced capacity to produce TMAO may contribute to their observed protection from CVD (Koeth et al., 2013).

Advanced Glycation End-Products

Endogenous formation of advanced glycation end-products (AGEs) is a normal consequence of metabolism, but may be up-regulated in conditions of metabolic stress (Davis, Prasad, Vijayagopal, Juma & Imrhan, 2014). They are also absorbed by the body from exogenous sources such as cigarette smoke and consumption of highly heated processed foods (Kellow & Savige, 2013). About 10% of AGEs ingested are absorbed, and of these two-thirds are deposited in the body tissues and one third is excreted by the kidneys (van Puyvelde, Mets, Njemini, Beyer & Bautmans, 2014). Advanced glycation end-products

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

are found in high concentrations in high-fat foods and meats cooked at high temperatures, with more formed by dry heat (grilling, frying, roasting, baking, barbecuing). They occur after a reaction between carbonyl groups on sugars and amine groups on proteins, and DNA and lipoproteins to form glycated proteins. This is the first step of the Maillard reaction, which is also responsible for browning reactions between carbohydrates and amines in cooked foods.

AGEs have been shown to contribute to chronic inflammation and oxidative stress, as well as dysfunction and destruction of pancreatic beta cells and insulin resistance, all of which are linked to chronic diseases such as cardiovascular disease and T2D (Jiao et al., 2013; Uribarri et al., 2010; White & Collinson, 2013; Kellow & Savige, 2013). Recent findings from the NIH-AARP Diet and Health Study showed an increased risk of pancreatic cancer in men in the highest quintile of AGE consumption compared to the lowest quintile (HR 1.43, 95% CI 1.06-1.93) (Jiao et al., 2015). Red meat intake was also associated with pancreatic cancer risk (HR 1.35, 95% CI 1.07-1.70) comparing the highest and lowest quintiles and this was attenuated after adjustment for AGE consumption, suggesting that dietary AGEs may partially explain the relationship between red meat intake and pancreatic cancer risk. In animal studies low-AGE diets reverse insulin resistance and chronic inflammation, inhibit the progression of atherosclerosis and prevent experimental diabetic nephropathy and neuropathy (Uribarri et al., 2007), but currently there is insufficient evidence in humans to promote the use of AGE-restricted diets (Kellow & Savige, 2013).

Animal derived foods that are high in fat and protein (particularly red and processed meats) are generally AGE-rich and prone to formation of AGEs during cooking. For example, the AGE content of beef increases from 707 kU/100 g when raw to 6071kU/100g when roasted and 10 058kU/100g when pan-fried. Processed meats, such as bacon and frankfurters have particularly high levels. By contrast, plant foods including fruits, vegetables, grains and legumes, have low levels of AGEs unless prepared with added animal fats such as butter (Uribarri et al., 2010). Guidelines to reduce AGE intake are consistent with established dietary guidelines for disease prevention. Table 4 compares the AGE content of a variety of red and processed meats versus other protein sources (Uribarri et al., 2010).

Neu5Gc

N-glycolylneuraminic acid (Neu5Gc) is an abundant sialic acid found mainly in red meat and to a smaller extent cow's milk but not in plant foods and poultry and only in trace amounts in fish (Tangvoranuntakul et al., 2003; Chen, Pan, Liu, Troy & Wang, 2014). Humans lack the enzyme responsible for its formation, however it can be metabolically incorporated into human tissues from dietary sources (Samraj, Laubli, Varki & Varki, 2014). It has been proposed that once ingested, the interaction between this 'foreign' molecule and anti-Neu5Gc antibodies results in chronic inflammation that promotes carcinogenesis and atherogenesis (Samraj et al., 2014). While further research is needed this could be one explanation for the association between red meat and chronic disease risk that isn't seen with other protein sources.

SOLUTIONS AND RECOMMENDATIONS

Meat Free and Low Meat Diets and Health

Several well-recognised dietary patterns have repeatedly been associated with a reduction in risk of chronic disease including vegetarian and vegan diets, Mediterranean diets, the Dietary Approaches to

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 4. AGE content of red and processed meats versus other protein sources

Food Item—Meat and Meat Substitutes	AGE kU/100g	Serving Size (g)	AGE kU/Serving
Beef, frankfurter, broiled	11270	90	10143
Beef, steak, pan fried with olive oil	10058	90	9052
Beef, steak, strips, stir fried with canola oil	9522	90	8570
Beef, steak, broiled	7479	90	6731
Beef, roast	6071	90	5464
Beef, hamburger (McDonald's Corp)	5418	90	4876
Beef, steak, microwaved	2687	90	2418
Beef, raw	707	90	636
Bacon, fried 5 mins, no oil	91577	13	11905
Bacon, microwaved, 2 slices, 3 mins	9023	13	1173
Sausage, beef and pork, pan fried	5426	90	4883
Pork, chop, pan fried	4752	90	4277
Ham, deli, smoked	2349	90	2114
Lamb, leg, broiled, 30 mins	2431	90	2188
Lamb, leg, microwave, 5 min	1029	90	926
Lamb, leg, raw	826	90	743
Chicken breast, breaded, deep fried	9722	90	8750
Chicken, back/thigh, roasted then BBQ	8802	90	7922
Chicken, breast, pan fried, 13 min, high	4938	90	4444
Chicken breast, skinless, poached	1076	90	968
Chicken, breast, skinless, raw	769	90	692
Salmon, broiled with olive oil	4334	90	3901
Salmon, Atlantic, frozen, poached, 7 mins	1801	90	1621
Tuna, fresh, baked, 25 mins	919	90	827
Salmon fillet, microwaved	912	90	821
Tuna, canned, chunk light, with water	452	90	407
Tofu, sautéed, outside	5877	90	5289
Tofu, broiled	4107	90	3696
Tofu, raw	788	90	709
Soy burger, skillet, with olive oil	437	30	131
Eggs, fried, one large	2749	45	1237
Egg, omelette, pan, low heat, 12 mins	223	30	67
Egg, scrambled, pan, high, olive oil, 1 min	243	30	73
Beans, red kidney, cooked 1 hr	298	100	298
Beans, red kidney, canned	191	100	191
Beans, red kidney, raw	116	100	116

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Stop Hypertension (DASH), and the Portfolio diet. These eating patterns are all rich in plant foods and contain little or no meat.

Vegetarian and Vegan Diets

A vegetarian diet consists mostly of plant-based foods including fruit, vegetables, legumes, nuts, seeds and grains. A lacto-ovo vegetarian diet also includes eggs and dairy products, while vegan diets contain no animal products. The complete absence of red meat is a distinguishing factor in both vegan and vegetarian diets.

There is now a significant amount of research to demonstrate the health benefits of vegetarian and vegan diets, particularly with respect to reducing chronic disease risk, including risk of CVD, T2D, MetS, hypertension and obesity. While vegetarians generally have a lower BMI (Tonstad et al., 2009), and tend to be more health conscious than non-vegetarians (Bedford & Barr, 2005), health outcomes remain better even when these factors are taken into account. Furthermore, a number of studies have shown increased longevity among vegetarians (Singh et al., 2003; Fraser & Shavlik, 2001). When compared to lacto-ovo-vegetarian diets, vegan diets appear to offer additional protection for obesity, hypertension, TD2, and cardiovascular mortality (Le & Sabate, 2014).

One of the proposed mechanisms for the protective effect of a vegan diet is the fibre intake and the impact on the intestinal microbial profile. The high fibre intake of a vegan diet may be a pivotal factor on gut microbiota and key measures of inflammation. In addition, the impact of plant *versus* animal protein on microbiota in the gut may provide further insight into variance in microbial profiles, MetS and inflammation variations in vegan and vegetarian diets (Glick-Bauer & Yeh, 2014).

The most consistent evidence for the health benefits of a vegetarian diet is for a reduced risk of CHD and CHD mortality. Several studies have demonstrated a lower incidence of CHD (Appleby, Thorogood, Mann & Key, 1999; Burr & Butland, 1988; Chang-Claude, Hermann, Eilber & Steindorf, 2005; Fraser, 1999; Snowdon, Phillips & Fraser, 1984) and an earlier combined analysis of 5 prospective studies reported a 24% lower risk of mortality from ischaemic heart disease (IHD) in vegetarians compared to meat-eaters, with lacto-ovo vegetarians having a 34% reduced risk and vegans a 26% reduced risk (Key et al., 1998, 1999). The benefit was apparent in those who had followed their diet for at least 5 years and was greater in younger age groups. A more recent meta-analysis of 7 studies involving 124 706 participants found that compared to non-vegetarians, vegetarians had a 9% lower risk of all-cause mortality (RR = 0.91; 95% CI, 66–1.16) and a 29% lower risk of mortality from IHD (RR = 0.71; 95% CI, 0.56–0.87) (Huang et al., 2012).

These findings are perhaps not surprising considering that vegetarians generally have a better cardiovascular risk profile than non-vegetarians with lower total and low density lipoprotein (LDL) cholesterol levels (Appleby, Thorogood, McPherson & Mann, 1995; Burr & Butland, 1988; West & Hayes, 1968; de Biase, Fernandes, Gianini & Duarte, 2007; Padler-Karavani et al., 2008; Nieman et al., 1989) a lower body weight (Rosell et al., 2006; Spencer, Appleby, Davey & Key, 2003; Tonstad et al., 2009), and a lower incidence of diabetes (Snowdon & Phillips, 1985; Tonstad et al., 2009; Vang, Singh, Lee, Haddad & Brinegar, 2008) and hypertension (Appleby, Davey & Key, 2002; Fraser, 2009; Rouse, Armstrong & Beilin, 1983), all of which contribute to CVD risk. A vegetarian diet may also reduce the susceptibility of LDL to oxidation (Lu et al., 2000).

Evidence suggests that a vegetarian diet may help in both the prevention and management of hypertension (Berkow & Barnard, 2005). A meta-analysis of 7 controlled clinical trials and 32 observational

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

studies reviewed the impact of vegetarian diets on blood pressure (BP). While there was only a small number of participants in the controlled trials (311 subjects), consumption of vegetarian diets was associated with a mean systolic BP (-4.8 mm Hg; 95% CI, -6.6 to -3.1) compared with the consumption of omnivorous diets. In the observational studies (a total of 21,604 subjects), vegetarian diets were associated with lower mean systolic BP (-6.9 mm Hg; 95% CI, -9.1 to -4.7) and diastolic BP (-4.7 mm Hg; 95% CI, -6.3 to -3.1) compared with omnivorous diets (Yokoyama et al., 2014). In the EPIC-Oxford study non-meat eaters had a lower prevalence of hypertension and lower systolic and diastolic blood pressures than meat eaters although this was largely due to differences in BMI (Appleby et al., 2002). The age-adjusted prevalence of self-reported hypertension was significantly different between the four diet groups, ranging from 15.0% in male meat eaters to 5.8% in male vegans, and from 12.1% in female meat eaters to 7.7% in female vegans (Appleby et al., 2002). Fish eaters and vegetarians had a similar, intermediate prevalence of hypertension. In the Adventist Health Study 2 (AHS-2), a significant, graded association with blood pressure was also observed (Fraser, 2009). Compared with non-vegetarians, vegans were 75% less likely to be treated for hypertension, lacto-ovo vegetarians 65% less likely, pesco-vegetarians 38% less likely and semi-vegetarians 23% less likely.

Vegetarian and vegan diets also offer significant benefits for the prevention and management of diabetes. Observational studies have demonstrated a significantly lower risk of T2D in individuals following a vegetarian diet compared to non-vegetarians (Snowdon & Phillips, 1985; Tonstad et al., 2009). The original AHS, involving 25 698 Seventh-day Adventist church members who were followed up for 21 years from recruitment in 1960, found that when compared to vegetarians, non-vegetarian women had a 1.4 times higher risk and non-vegetarian men a 1.8 times higher risk of developing T2D (Snowdon & Phillips, 1985). In the more recent AHS-2 a cross-sectional analysis of baseline data from 60 903 Seventh-day Adventist Church members from North America found that those following a vegan diet had a diabetes prevalence approximately one-third that of non-vegetarians (2.9% versus 7.6%), while the lacto-ovo vegetarians, pesco-vegetarians and semi-vegetarians had an intermediate diabetes prevalence of 3.2%, 4.8% and 6.1% respectively (Tonstad et al., 2009). After adjusting for confounding factors, those following a vegan diet had almost a 50% reduction in the risk of developing T2D compared to non-vegetarians. The risk reduction became incrementally smaller as more animal products were consumed - lacto-ovo vegetarians had a 46% reduction in risk, pesco-vegetarians a 30% reduction in risk and semi-vegetarians a 24% reduction in risk (Tonstad et al., 2009). The researchers also looked at the incidence of diabetes in a subset of 15 200 men and 26 187 women without diabetes at baseline. They were followed for 2 years and after adjustment for confounding factors, vegans were found to have a 62% lower risk of developing type 2 diabetes and lacto-ovo vegetarians a 38% lower risk compared to non-vegetarians (Tonstad et al., 2013).

The AHS-2 also demonstrated a reduced risk of MetS in vegetarians. A cross sectional analysis of 773 adults taking part in the study, just over half of whom were vegetarian (35%) or semi-vegetarian (16%), found the prevalence of MetS to be 23% in the vegetarians, 31% in semi-vegetarians and 37% in non-vegetarians (Rizzo, Sabate, Jaceldo-Siegl & Fraser, 2011). Compared to non-vegetarians, those following a vegetarian diet had 56% lower risk (adjusted OR 0.44, 95% CI 0.30-0.64) of having MetS.

There are only a few intervention studies assessing the effects of a vegetarian diet in people with diabetes, and weight loss has generally been greater on the vegetarian diets, making it difficult to determine the independent effect of the diets (Jenkins, Kendall, Marchie, Jenkins, et al., 2003). However, greater weight loss may be one of the advantages of a vegetarian diet as even small amounts of weight loss can help in both the prevention and management of T2D. One study comparing a low fat vegan diet with a

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

diet based on the American Diabetes Association (ADA) guidelines found that the vegan diet reduced HbA1c levels significantly more than the ADA diet (1.23 versus 0.38 percentage points in those who didn't change medication) (Barnard et al., 2006). Furthermore, 43% of subjects were able to reduce their medication compared to only 26% in the ADA group. Similarly, a study of 74 subjects with T2D randomised to a low vegetarian diet (animal foods restricted to 1 low fat yoghurt/day) or a control diet (both isocaloric and calorie restricted) demonstrated significantly greater weight loss, reductions in waist circumference, visceral and subcutaneous fat, oxidative stress markers and adipokines, and improvements in insulin sensitivity on the vegetarian diet compared to the control diet (Kahleova et al., 2011). Like the previous study, 43% of subjects were able to reduce their use of diabetes medication compared to only 5% of the control group. Earlier studies of low fat vegan diets in patients with type 2 diabetes have also demonstrated significant improvements in blood glucose control and blood fats, as well as considerable reductions in medication use (Barnard, Jung & Inkeles, 1994; Nicholson et al., 1999) although the larger of these studies was not controlled.

Overall, cancer rates in vegetarians appear to be moderately lower than others living in the same communities, and life expectancy appears to be greater (Fraser, 1999; Key et al., 2009; Fraser, 2009). A meta-analysis of 7 prospective cohort studies involving 124 706 participants showed an 18% reduction in incidence of all forms of cancer than non-vegetarians (RR = 0.82; 95% CI, 0.67–0.97) (Huang et al., 2012). Recent results from the AHS-2 found that among 69 120 participants, vegetarians were 8% less likely to develop cancer than non-vegetarians (HR, 0.92; 95% CI, 0.85–0.99) for both genders combined (Tantamango-Bartley, Jaceldo-Siegl, Fan & Fraser, 2013). A statistically significant association was found between vegetarian diets and cancers of the gastrointestinal tract (HR 0.76; 95% CI, 0.63–0.90), while vegan diets were associated with a significant reduction in risk of overall cancer (HR 0.84; 95% CI, 0.72–0.99) in both genders combined and for female-specific cancers (HR 0.66; 95% CI, 0.47–0.92) (Tantamango-Bartley et al., 2013).

Despite the current popularity of high protein, low carbohydrate diets for weight management, research has consistently shown that vegetarians, and particularly vegans, are leaner than their meat-eating counterparts (Berkow & Barnard, 2006; Rosell et al., 2006; Spencer et al., 2003; Tonstad et al., 2009). The EPIC-Oxford study compared weight gain over 5 years in almost 22 000 meat-eating, fish-eating, vegetarian, and vegan men and women and found that weight gain was lowest in the vegan group and those who, during follow-up, had changed to a diet containing fewer animal foods (Rosell et al., 2006). Similarly, the AHS-2 found that mean BMI was lowest in vegans (23.6 kg/m^2) and incrementally higher in lacto-ovo vegetarians (25.7 kg/m^2), pesco-vegetarians (26.3 kg/m^2), semi-vegetarians (27.3 kg/m^2), and non-vegetarians (28.8 kg/m^2) (Tonstad et al., 2009). A recent randomised control trial involving 50 subjects who were assigned to a vegan, vegetarian, pesco-vegetarian, semi-vegetarian or omnivorous diet with unrestricted energy intake for six months to investigate effects of plant-based diets on short-term weight loss found that those on the vegan diet lost significantly more weight than the other dietary patterns. (Turner-McGrievy, Davidson, Wingard & Billings, 2014). Together these studies suggest that plant-based diets may be a useful strategy to combat the obesity epidemic in western countries.

Mediterranean Diets

The most well researched dietary pattern is the Mediterranean diet (MD) which can be described as the diet followed in the olive growing areas of Mediterranean countries. The MD is characterized by a high intake of fruits, vegetables, cereals (mainly wholegrain varieties), nuts and seeds. Fish, seafood, poultry,

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

and eggs are consumed in moderation, while there is low consumption of dairy foods, red meat, processed meats and sweets (Trichopoulou & Lagiou, 1997). A key characteristic of the diet is the abundance of extra virgin olive oil (EVOO), which along with nuts provides the primary sources of fat in the diet. Foods included in the MD are minimally processed, rich in fibre, antioxidant polyphenols and essential micronutrients (Toledo et al., 2013).

A meta-analysis of 36 prospective cohort studies including a combined study population of over 4 172 412 subjects with a follow-up ranging from 4.9 to 16 years, confirmed the significant and consistent protective effects of the MD on chronic disease risk. A 2-point increase in adherence to the MD was associated with an 8% reduction in overall mortality (RR 0.92, 95% CI: 0.91-0.93), a 10% reduced risk of cardiovascular disease (RR 0.90, 95% CI: 0.87-0.92), and a 4% reduction in cancer incidence (RR 0.96, 95% CI 0.95-0.97) (Sofi, Macchi, Abbate, Gensini & Casini, 2014). A second recent meta-analysis found that a 2-point increase in adherence to a MD was associated with a 13% reduction in the risk of cardiovascular events (pooled RR 0.87, 95% CI 0.85-0.90 (Martinez-Gonzalez & Bes-Rastrollo, 2014). Similarly, a recent systematic review of 37 studies looking at the association between adherence to a MD and prevention of cardiodiabesity (incorporating T2D, obesity, MetS and CVD) found that 33 of these studies provided strong evidence for an association between adherence to a MD and reduced risk of collective cardiodiabesity risk (Garcia-Fernandez, Rico-Cabanas, Rosgaard, Estruch & Bach-Faig, 2014). An earlier meta-analysis of 50 studies (35 clinical trials, 2 prospective and 13 cross-sectional) found that adherence to a MD was associated with a reduced risk of MetS while the clinical studies showed a protective effect of the MD on individual components of the MetS including WC, HDL-cholesterol, TG, systolic and diastolic BP and glucose levels (Kastorini et al., 2011).

The benefits of a MD demonstrated in observational studies have been confirmed in the PREDIMED study, a large randomised controlled trial in individuals at high risk of cardiovascular disease, comparing the effects of a MD enriched with extra virgin olive oil (EVOO) or mixed nuts versus a low fat diet on the primary prevention of CVD. The primary outcome for the study was a reduced risk of CVD and the multivariable –adjusted hazard ratios for cardiovascular disease (a composite of myocardial infarction, stroke and death from cardiovascular causes) were 0.70 (95% CI, 0.54-0.92) and 0.72 (95% CI, 0.54-0.96) for the group assigned to a MD with added EVOO and nuts respectively (Estruch et al., 2013). On the basis of results of an interim analysis, the trial was stopped after a median follow-up of 4.8 years. Combining the results of the PREDIMED study with those of the Lyon Diet Heart Study (an earlier secondary prevention trial), showed that intervention with a MD was associated with a 38% relative reduction in the risk of CVD clinical events (Martinez-Gonzalez & Bes-Rastrollo, 2014).

The PREDIMED study has also shown benefits of a MD for diabetes prevention and management (Salas-Salvadó et al., 2011; Lasa et al., 2014) as well as a greater reversion of the MetS with a MD diet (Babio et al., 2014). In a subgroup of 418 non-diabetic subjects who were randomly assigned to one of the three dietary patterns the incidence rate of diabetes per 1000 person-years was 24.6 (MD plus EVOO), 26.8 (MD plus nuts) and 46.6 for the control group. After adjustment for several confounders, the risk of developing diabetes was reduced by more than 50% in the MD groups compared with the low-fat group (Salas-Salvadó et al., 2011). In a subgroup of 191 study participants with T2D, those following a MD supplemented with either EVOO or nuts, total body weight and glucose metabolism were improved to the same extent as the low-fat diet, despite higher energy intakes than the low fat diet group (Lasa et al., 2014). An analysis of results from 5801 participants in PREDIMED found that after 4.8 years of follow-up, a MD did not reduce the risk of developing MetS but those in the MD groups who

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

had MetS at baseline were more likely to experience reversion compared to the low fat diet group and had significant reductions in central obesity (Babio et al., 2014).

An early analysis of the PREDIMED study showed that after just 3 months, both MD groups had improved metabolic risk factors compared to the low fat diet group, including lower blood pressure, improved lipid profiles, decreased insulin resistance, and reduced concentrations of inflammatory molecules compared with the low-fat diet (Estruch et al., 2006). The mean changes in the MD with olive oil group and the MD with nuts group compared with the low-fat diet, were -0.39mmol/L and -0.30mmol/L, respectively, for plasma glucose levels, -5.9mmHg and -7.1mmHg, respectively for systolic blood pressure, and -0.38 and -0.26, respectively, for the cholesterol–HDL cholesterol ratio. The MD with olive oil also reduced C-reactive protein (CRP) levels by 0.54 mg/L compared with the low-fat diet (Estruch et al., 2006).

DASH Diet

The Dietary Approaches to Stopping Hypertension (DASH) diet is an eating pattern that is low in saturated fat, cholesterol and total fat with an increased intake of fruits, vegetables, and fat-free or low-fat dairy products. The diet is high in dietary fibre and minerals including potassium, magnesium and calcium. It focuses on fruits, vegetables, whole grains, fish, poultry, and nuts and legumes but restricts intake of red meat, sweets, added sugars, and sugar-sweetened drinks.

The original DASH trial found that this dietary intervention can help to prevent hypertension, may be as effective as medication for treating mild hypertension and might prevent or delay the need for medication in people with borderline high blood pressure (Appel et al., 1997). After 8 weeks the DASH diet lowered systolic blood pressure by 5.5/3.0mmHg more than a control diet and to a greater extent than a diet rich in fruit and vegetables. A greater reduction (11.4/5.5mmHg) was observed in those with baseline hypertension. Another study combining the DASH diet with a reduction in sodium intake found that the blood pressure lowering effect of the DASH diet is additive to a reduction in sodium intake (Bray et al., 2004).

A meta-analysis of observational studies has demonstrated that DASH-style dietary interventions can significantly reduce the risk of CVD, CHD, stroke and HF by 20%, 21% 19% and 29% respectively (Salehi-Abargouei, Maghsoudi, Shirani & Azadbakht, 2013). A systematic review of the DASH diet and risk of T2D found that a DASH diet can reduce fasting insulin levels but not fasting glucose levels or measures of insulin resistance (Shirani, Salehi-Abargouei & Azadbakht, 2013). In a combined analysis of the NHS and HPFS, adherence to a DASH-style diet was associated with a lower risk of CRC, with a 20% reduction in risk for those in the top versus bottom quintiles of DASH diet score (Fung, Hu, et al., 2010). Similarly, in the NHS, the DASH diet was associated with a 20% lower risk of estrogen receptor-negative breast cancer in postmenopausal women, comparing the highest and lowest quintiles (Fung, Hu, Hankinson, Willett & Holmes, 2011).

Portfolio Diet

The Portfolio diet was designed by Canadian researchers to combine a number of foods known to have cholesterol-lowering properties – plant sterols (from plant sterol enriched foods), soy protein (from foods such as soy milk and tofu), soluble fibres (from oats, barley, psyllium and certain vegetables such

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

as okra and eggplant), and almonds (a handful each day). The diet is also limited in animal foods (meat and dairy products) and encourages the consumption of 5 to 9 serves of fruit and vegetables each day.

In the original study of the Portfolio diet, 34 hyperlipidaemic subjects all underwent three 1-month treatments in random order as outpatients. Treatment arms included a very low saturated fat diet (control diet), the same control diet with the inclusion of 20mg of lovastatin (statin diet) and the intervention 'portfolio' diet as described above. LDL concentrations decreased by 8.5%, 33.3% and 29.6% in the control, statin and portfolio diets respectively, demonstrating that the diet is as effective as a low dose of statin and can achieve a cholesterol lowering effect more than three times that of a low saturated fat diet over 4 weeks (Jenkins et al., 2002). Several subsequent studies have shown similar benefits (Gigleux et al., 2007; Jenkins, Kendall, Marchie, Faulkner, et al., 2003). The Portfolio diet was also as effective as a first-generation statin in reducing CRP (Jenkins et al., 2005) and has been shown to lower blood pressure, although the latter was associated with the almond content of the diet (Jenkins, Kendall, Nguyen, et al., 2008)

A longer term study of 55 subjects found that after following the Portfolio diet for 3 months, subjects had an average reduction in LDL cholesterol levels of 14.0% and this was maintained at 12.8% at the end of the 12 month study. Almost one-third of participants had reductions in LDL cholesterol of more than 20% after 12 months (Jenkins, Kendall, Faulkner, et al., 2008). Again, these results were similar to that achieved by taking a first generation statin.

Implications and Recommendations for Adopting a Low Meat or Meat Free Diet

While not everyone will choose to adopt a meat-free diet, any reduction in meat intake is likely to benefit health, particularly reducing chronic disease risk. This is evident from the findings of the AHS-2 which, in a population of more than 73 000 men and women, compared health outcomes of non-vegetarians, semi-vegetarians (those who ate meat, poultry or fish less than 1/week), pesco-vegetarians (those who ate fish but no other animal flesh), lacto-ovo vegetarians (those who ate eggs and dairy products but no animal flesh), and vegans (those who ate no animal products). They found that BMI, and risk of diabetes and hypertension (even after controlling for weight) decreased incrementally from the non-vegetarians to vegans, as less animal foods were included in the diet (Orlich et al., 2013).

The World Cancer Research Fund, as part of their cancer prevention recommendations, advise limiting the consumption of red meats (such as beef, pork and lamb) and avoiding processed meats. They recommend the population average consumption of red meat to be no more than 300 g (11 oz) a week, very little if any of which should be processed, and advise individuals who eat red meat to consume less than 500 g (18 oz) a week and to avoid or minimise processed meat (WCRF, 2007). The Australian Dietary Guidelines similarly recommend limiting red meat intake (to 455g per week of cooked lean red meat) on the basis of research showing regular consumption of large quantities of red meat (100–120g/day) may be associated with an increased risk of colorectal cancer (NHMRC, 2013). At the time of writing, the US Dietary Guidelines Advisory Committee had just released their recommendations to the U.S. Department of Agriculture (USDA) and the Department of Health and Human Services (USDHHS), recommending a reduction in red and processed meat consumption to benefit both health and the environment (USDHHS, 2015).

Three studies looking at the impact of decreasing meat intake to reduce greenhouse gas (GHG) emissions have shown such reductions in intake would also have significant population health benefits. Friel et al. modelled the potential benefits of a 30% reduction in livestock production (to achieve the UK

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Committee on Climate Change recommended targets) on the burden of IHD (Friel et al., 2009). They estimated that IHD burden would decrease by about 15% in the UK and 16% in Sao Paulo city. Aston et al. conducted a modelling study using dietary intake data from the National Diet and Nutrition Survey of British Adults and found that if there was a doubling of the number of vegetarians (to 4.7% of men and 12.3% of women) and the remainder of the population adopted the current level of meat intake of the lowest quintile of non-vegetarians, there would be a significant reduction in risk of CHD, diabetes and CRC (Aston, Smith & Powles, 2012). Scarborough, while not looking independently at the impact of reduced red meat consumption, estimated that reducing the consumption of meat and dairy foods by 50% and replacing them with fruits, vegetables and cereals, would not only significantly reduce GHG emissions but would result in close to 37 000 deaths delayed or averted per year in the UK (Scarborough, Allender, Clarke, Wickramasinghe & Rayner, 2012). Together these studies show that reducing red meat intake could have significant benefits for both the environment and reducing chronic disease risk.

The main concerns cited in regards to adopting a vegetarian diet are acceptability and nutritional concerns. However several studies have found that low-fat vegetarian or vegan diets were well accepted, and similar to that of standard low fat diets (Barnard, Scialli, Turner-Mcgrievy, & Lanou 2004; Franklin, Kolasa, Griffin, Mayo & Badenhop, 1995; Barnard et al., 2009; Berkow, Bernard, Katcher & Eckart, 2010). For example Barnard et al., in a study of subjects with T2D, found that a vegan diet was equally as acceptable as a more conventional (American Diabetes Association) diet among study participants (Barnard et al., 2009). In fact, after 22 weeks, significantly more of the vegan diet group (67%) met group-specific diet adherence criteria compared to the ADA diet group (44%) (Barnard et al., 2009). In a recent weight loss trial in which overweight subjects were randomised to one of five low fat low GI diets (vegan, vegetarian, pesco-vegetarian, semi-vegetarian and omnivorous) dietary adherence was similar in all groups at two and six months (Turner-McGrievy, Davidson, Wingard, Wilcox & Frongillo, 2015). The introduction of vegetarian meals in schools and workplaces has also shown to be well accepted (de Keyzer et al., 2012; Katcher, Ferdowsian, Hoover, Cohen & Barnard, 2010).

When it comes to nutrition, the American Dietetic Association supports the fact that, in addition to their health benefits, well planned vegetarian diets, including vegan diets, are nutritionally adequate and are appropriate for individuals during all stages of life (Craig & Mangels, 2009). Some nutrients can be more difficult to obtain on a vegetarian diet, but careful planning and in some cases the use of fortified foods or supplements, can ensure that an individual's nutrition needs are met while maximising the health benefits of a vegetarian or vegan diet (Reid, Marsh, Zeuschner, Saunders & Baines, 2013).

A cross-sectional analysis of the US National Health and Nutrition Examination Survey (NHANES) (1999-2004) data looking at diet quality found that lacto-ovo vegetarian diets are nutrient dense and consistent with dietary guidelines. In fact vegetarians had higher mean intakes of fibre, vitamins A, C, and E, thiamin, riboflavin, folate, calcium, magnesium, and iron, and lower intakes of total fat, saturated fat and cholesterol (Farmer, Larson, Fulgoni, Rainville & Liepa, 2011). Overall diet quality, measured by the US Department of Agriculture's Healthy Eating Index 2005 was no different between vegetarians and non-vegetarians. A survey of US vegetarians showed that those who define themselves as vegetarian have dietary patterns that are generally healthier than non-vegetarians, being lower in total fat, saturated fat, and cholesterol and higher in fibre, with a higher intake of grains, legumes, vegetables (green leafy and yellow) and fruit (Haddad & Tanzman, 2003). Vitamin C, vitamin E, thiamine, folate, calcium and magnesium intakes were also higher than in those who ate meat, however iron, zinc and vitamin B12 intakes were lower.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

In the study by Barnard et al., both the vegan and ADA dietary groups reduced their intake of energy, protein, fat, cholesterol and sodium. However, the vegan diet group had a greater reduction in fat and cholesterol intakes (Barnard et al., 2009; Turner-McGrievy et al., 2008). The vegan dieters significantly increased their intake of carbohydrate, dietary fibre, vitamins A, C and K, beta-carotene, folate, magnesium and potassium, but reduced their intakes of calcium and vitamin B12 while the ADA diet group reduced their intake of carbohydrate and iron but reported no significant increases in nutrient intake (Turner-McGrievy et al., 2008). The vegan diet group was also found to have a significant increase in their Alternate Healthy Eating Index (AHEI) score while the ADA diet group did not, and there was a negative correlation between AHEI scores and HbA1c levels (Fung, Mc-Cullough, van Dam & Hu, 2007; Fung et al., 2005; McCullough et al., 2002). The AHEI is a nine-component dietary index used to rate foods and nutrients related to chronic disease risk, and previous research has linked a higher AHEI score with a lower risk of type 2 diabetes and cardiovascular disease (Turner-McGrievy et al., 2008). Similarly, in a weight loss trial comparing diets containing different amounts of animal foods (vegan, vegetarian, pesco-vegetarian, semi-vegetarian and omnivorous) the vegan diet group had the greatest reduction in saturated fat and cholesterol and a greater increase in fibre intake compared to the pesco-vegetarian, semi-vegetarian and omnivorous dieters (Turner-McGrievy et al., 2015).

'Meat-Free Mondays'

One very successful campaign to encourage a population-wide reduction in meat intake is the Meatless Monday movement (http://www.meatlessmonday.com) or Meat-Free Mondays (http://www.meatfreemondays.com/). The campaign began in the US in 2003 in association with the Johns Hopkins Bloomberg School of Public Health's Center for a Liveable Future and is now a global movement, active in 36 countries including Denmark, France, Germany, Italy, Japan, the United States, Britain, Canada, Israel and Australia, and still growing. The Meat-Free Mondays campaign encourages individuals to give up eating meat one day per week, to benefit both their health and the environment. This encourages participants to learn to enjoy plant-based meals without necessarily becoming vegetarian.

FUTURE RESEARCH DIRECTIONS

Considering the increased interest in low carbohydrate high protein (LCHP) diets for weight management, well-controlled longer term studies comparing such diets with plant-based diets such as vegetarian, vegan and Mediterranean diets are warranted. These studies should investigate the effects of these diets not only on common metabolic and cardiovascular risk factors but also on non-traditional risk factors reflective of chronic disease risk. Animal studies have shown deleterious effects of low carbohydrate high protein diets on metabolic and cardiovascular function. In one study mice fed a LCHP diet developed adverse vascular effects including increased aortic atherosclerosis and an impaired ability to generate new vessels in response to tissue ischemia, independent of traditional risk factors including cholesterol levels, inflammatory markers and markers of oxidative stress (Foo et al., 2009). A second study found that adult (but not younger) mice fed a LCHP diet developed marked glucose intolerance and arterial dysfunction resulting in cardiovascular changes similar to those seen in old age (National Institute of Health, 2014). The same authors similarly found adverse effects on glucose tolerance and major arterial dysfunction in older mice fed a LCHP diet (Baron et al., 2014). Together with the wealth of information

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

showing health benefits, reduced chronic disease risk and lower mortality with plant-based diets, these studies highlight the need to consider the longer-term health effects of diets high in animal protein and fat.

CONCLUSION

There is now a significant body of evidence demonstrating an association between red and processed meat intakes and chronic disease risk, and a number of plausible mechanisms to explain this association. On the other hand, low meat or meat-free diets are associated with improve health outcomes and reduced disease risk. In Western countries in particular, plant-based diets low in red meat may present a significant advantage over meat-based diets, not only for the environment but also for reducing chronic disease risk and increasing longevity.

REFERENCES

Abete, I., Romaguera, D., Vieira, A. R., Lopez de Munain, A., & Norat, T. (2014). Association between total, processed, red and white meat consumption and all-cause, CVD and IHD mortality: A meta-analysis of cohort studies. *The British Journal of Nutrition*, *112*(5), 762–775.

Abid, Z., Cross, A. J., & Sinha, R. (2014). Meat, dairy, and cancer. *The American Journal of Clinical Nutrition*, *100*(Supplement 1), 386S–393S.

Alexander, D. D., & Cushing, C. A. (2009). Quantitative assessment of red meat or processed meat consumption and kidney cancer. *Cancer Detection and Prevention*, *32*(5-6), 340–351.

Alexander, D. D., Mink, P. J., Cushing, C. A., & Sceurman, B. (2010). A review and meta-analysis of prospective studies of red and processed meat intake and prostate cancer. *Nutrition Journal*, *9*(50), 1–17.

Alexander, D. D., Morimoto, L. M., Mink, P. J., & Cushing, C. A. (2010). A review and meta-analysis of red and processed meat consumption and breast cancer. *Nutrition Research Reviews*, *23*(2), 349–365.

Altorf-van der Kuil, W., Engberink, M. F., Vedder, M. M., Boer, J. M., Verschuren, W. M., & Geleijnse, J. M. (2012). Sources of dietary protein in relation to blood pressure in a general dutch population. *Public Library of Science (Plos). One*, *7*(2), 1–8.

Andrews, N. C. (1999). Disorders of iron metabolism. *The New England Journal of Medicine*, *341*(26), 1986–1995.

Appel, L. J., Moore, T. J., Obarzanek, E., Vollmer, W. M., Svetkey, L. P., Sacks, F. M., & Karanja, N. et al. (1997). A clinical trial of the effects of dietary patterns on blood pressure. DASH Collaborative Research Group. *The New England Journal of Medicine*, *336*(16), 1117–1124.

Appleby, P. N., Davey, G. K., & Key, T. J. (2002). Hypertension and blood pressure among meat eaters, fish eaters, vegetarians and vegans in EPIC-Oxford. *Public Health Nutrition*, *5*(5), 645–654.

Appleby, P. N., Thorogood, M., Mann, J. I., & Key, T. J. (1999). The Oxford Vegetarian Study: An overview. *The American Journal of Clinical Nutrition*, *70*(3Suppl), 525S–531S.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Appleby, P. N., Thorogood, M., McPherson, K., & Mann, J. I. (1995). Associations between plasma lipid concentrations and dietary, lifestyle and physical factors in the Oxford Vegetarian Study. *Journal of Human Nutrition and Dietetics*, *8*(5), 305–314.

Ascherio, A., Katan, M. B., Zock, P. L., Stampfer, M. J., & Willett, W. C. (1999). Trans fatty acids and coronary heart disease. *The New England Journal of Medicine*, *340*(25), 1994–1998.

Aston, L. M., Smith, J. N., & Powles, J. W. (2012). Impact of a reduced red and processed meat dietary pattern on disease risks and greenhouse gas emissions in the UK: A modelling study. *British Medical Journal*, *2*(5), 1–9.

Astrup, A., Dyerberg, J., Elwood, P., Hermansen, K., Hu, F. B., Jakobsen, M. U., & Willett, W. C. et al. (2011). The role of reducing intakes of saturated fat in the prevention of cardiovascular disease: Where does the evidence stand in 2010? *The American Journal of Clinical Nutrition*, *93*(4), 684–688.

Aune, D., Chan, D. S., Vieira, A. R., Navarro Rosenblatt, D. A., Vieira, R., Greenwood, D. C., & Norat, T. et al. (2013). Red and processed meat intake and risk of colorectal adenomas: A systematic review and meta-analysis of epidemiological studies. *Cancer Causes & Control*, *24*(4), 611–627.

Aune, D., Ursin, G., & Veierod, M. B. (2009). Meat consumption and the risk of type 2 diabetes: A systematic review and meta-analysis of cohort studies. *Diabetologia*, *52*(11), 2277–2287.

Azadbakht, L., & Esmaillzadeh, A. (2009). Red meat intake is associated with metabolic syndrome and the plasma C-reactive protein concentration in women. *The Journal of Nutrition*, *139*(2), 335–339.

Babio, N., Sorli, M., Bullo, M., Basora, J., Ibarrola-Jurado, N., Fernandez-Ballart, J., & Salas-Salvado, J. et al. (2012). Association between red meat consumption and metabolic syndrome in a Mediterranean population at high cardiovascular risk: Cross-sectional and 1-year follow-up assessment. *Nutrition, Metabolism, and Cardiovascular Diseases*, *22*(3), 200–207.

Babio, N., Toledo, E., Estruch, R., Ros, E., Martinez-Gonzalez, M. A., Castaner, O., & Salas-Salvado, J. et al. (2014). Mediterranean diets and metabolic syndrome status in the PREDIMED randomized trial. *Canadian Medical Association Journal*, *186*(17), E649–E657.

Baer, D. J., Judd, J. T., Clevidence, B. A., & Tracy, R. P. (2004). Dietary fatty acids affect plasma markers of inflammation in healthy men fed controlled diets: A randomized crossover study. *The American Journal of Clinical Nutrition*, *79*(6), 969–973.

Bandera, E. V., Kushi, L. H., Moore, D. F., Gifkins, D. M., & McCullough, M. L. (2007). Consumption of animal foods and endometrial cancer risk: A systematic literature review and meta-analysis. *Cancer Causes & Control*, *18*(9), 967–988.

Bao, W., Bowers, K., Tobias, D. K., Hu, F. B., & Zhang, C. (2013). Prepregnancy dietary protein intake, major dietary protein sources, and the risk of gestational diabetes mellitus: A prospective cohort study. *Diabetes Care*, *36*(7), 2001–2008.

Bao, W., Rong, Y., Rong, S., & Liu, L. (2012). Dietary iron intake, body iron stores, and the risk of type 2 diabetes: A systematic review and meta-analysis. *BioMed Central Medicine*, *10*, 1–13.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Barnard, N. D., Cohen, J., Jenkins, D. J., Turner-McGrievy, G., Gloede, L., Jaster, B., & Talpers, S. et al. (2006). A low-fat vegan diet improves glycemic control and cardiovascular risk factors in a randomized clinical trial in individuals with type 2 diabetes. *Diabetes Care*, *29*(8), 1777–1783.

Barnard, N. D., Gloede, L., Cohen, J., Jenkins, D. J., Turner-McGrievy, G., Green, A. A., & Ferdowsian, H. (2009). A low-fat vegan diet elicits greater macronutrient changes, but is comparable in adherence and acceptability, compared with a more conventional diabetes diet among individuals with type 2 diabetes. *Journal of the American Dietetic Association*, *109*(2), 263–272.

Barnard, N. D., Scialli, A. R., Turner-Mcgrievy, G., & Lanou, A. J. (2004). Acceptability of a low-fat vegan diet compares favourably to a step II diet in a randomized, controlled trial. *Journal of Cardiopulmonary Rehabilitation*, *24*(4), 229–235.

Baron, S., Bedarida, T., Cottart, C. H., Vilbert, F., Vessieres, E., Ayer, A.,… Nivet-Antoine, V. (2014). Dual effects of resveratrol on arterial damage induced by insulin resistance in aged mice. *The Journal of Gerontology: Series A Biological Sciences and Medical Sciences, 69*(3), 260-269. doi: 2.10.1093/gerona/glt1081

Barnard, R. J., Jung, T., & Inkeles, S. B. (1994). Diet and exercise in the treatment of NIDDM. The need for early emphasis. *Diabetes Care*, *17*(12), 1469–1472.

Bastide, N. M., Pierre, F. H., & Corpet, D. E. (2011). Heme iron from meat and risk of colorectal cancer: A meta-analysis and a review of the mechanisms involved. *Cancer Prevention Research (Philadelphia, Pa.)*, *4*(2), 177–184.

Beasley, J. M., & Wylie-Rosett, J. (2013). The role of dietary proteins among persons with diabetes. *Current Atherosclerosis Reports*, *15*(9), 1–8.

Bedford, J. L., & Barr, S. I. (2005). Diets and selected lifestyle practices of self-defined adult vegetarians from a population-based sample suggest they are more 'health conscious'. *The International Journal of Behavioral Nutrition and Physical Activity*, *2*(1), 1–11.

Bennett, B. J., de Aguiar Vallim, T. Q., Wang, Z., Shih, D. M., Meng, Y., Gregory, J., & Lusis, A. J. et al. (2013). Trimethylamine-N-oxide, a metabolite associated with atherosclerosis, exhibits complex genetic and dietary regulation. *Journal of Cell Metabolism*, *17*(1), 49–60.

Berkow, S. E., & Barnard, N. (2006). Vegetarian diets and weight status. *Nutrition Reviews*, *64*(4), 175–188.

Berkow, S. E., & Barnard, N. D. (2005). Blood pressure regulation and vegetarian diets. *Nutrition Reviews*, *63*(1), 1–8.

Berkow, S. E., Bernard, N., Katcher, H., & Eckart, J. (2010). Four Theraputic Diets: Adherance and Acceptanility. *Canadian Journal of Dietetic Practice and Researc*, *71*(4), 199–204.

Bernstein, A. M., Pan, A., Rexrode, K. M., Stampfer, M., Hu, F. B., Mozaffarian, D., & Willett, W. C. (2012). Dietary protein sources and the risk of stroke in men and women. *Stroke*, *43*(3), 637–644.

Bernstein, A. M., Sun, Q., Hu, F. B., Stampfer, M. J., Manson, J. E., & Willett, W. C. (2010). Major dietary protein sources and risk of coronary heart disease in women. *Circulation*, *122*(9), 876–883.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Bonequi, P., Meneses-Gonzalez, F., Correa, P., Rabkin, C. S., & Camargo, M. C. (2013). Risk factors for gastric cancer in Latin America: A meta-analysis. *Cancer Causes & Control*, *24*(2), 217–231.

Bowers, K., Yeung, E., Williams, M. A., Qi, L., Tobias, D. K., Hu, F. B., & Zhang, C. (2011). A prospective study of prepregnancy dietary iron intake and risk for gestational diabetes mellitus. *Diabetes Care*, *34*(7), 1557–1563.

Bray, G. A., Vollmer, W. M., Sacks, F. M., Obarzanek, E., Svetkey, L. P., & Appel, L. J. (2004). A further subgroup analysis of the effects of the DASH diet and three dietary sodium levels on blood pressure: Results of the DASH-Sodium Trial. *The American Journal of Cardiology*, *94*(2), 222–227.

Broedbaek, K., Siersma, V., Andersen, J. T., Petersen, M., Afzal, S., & Hjelvang, B., cPoulsen, H.E. (2011). The association between low-grade inflammation, iron status and nucleic acid oxidation in the elderly. *Free Radical Research*, *45*(4), 409–416.

Brugère, J. F., Borrel, G., Gaci, N., Tottey, W., O'Toole, P. W., & Malpuech-Brugere, C. (2014). Archaebiotics: Proposed therapeutic use of archaea to prevent trimethylaminuria and cardiovascular disease. *Gut Microbes*, *5*(1), 5–10.

Burr, M. L., & Butland, B. K. (1988). Heart disease in British vegetarians. *The American Journal of Clinical Nutrition*, *48*(3Suppl), 830–832.

Chan, D. S., Lau, R., Aune, D., Vieira, R., Greenwood, D. C., Kampman, E., & Norat, T. (2011). Red and processed meat and colorectal cancer incidence: Meta-analysis of prospective studies. *Public Library of Science (PLoS). One*, *6*(6), 1–11.

Chang-Claude, J., Hermann, S., Eilber, U., & Steindorf, K. (2005). Lifestyle determinants and mortality in German vegetarians and health-conscious persons: Results of a 21-year follow-up. *Cancer Epidemiology, Biomarkers & Prevention*, *14*(4), 963–968.

Chen, G. C., Lv, D. B., Pang, Z., & Liu, Q. F. (2013). Red and processed meat consumption and risk of stroke: A meta-analysis of prospective cohort studies. *European Journal of Clinical Nutrition*, *67*(1), 91–95.

Chen, Y., Pan, L., Liu, N., Troy, F. A., & Wang, B. (2014). LC-MS/MS quantification of N-acetylneuraminic acid, N-glycolylneuraminic acid and ketodeoxynonulosonic acid levels in the urine and potential relationship with dietary sialic acid intake and disease in 3- to 5-year-old children. *The British Journal of Nutrition*, *111*(2), 332–341.

Chen, Y. H., Xia, E. Q., Xu, X. R., Li, S., Ling, W. H., Wu, S., & Li, H. B. et al. (2012). Evaluation of benzo[a]pyrene in food from China by high-performance liquid chromatography-fluorescence detection. *International Journal of Environmental Research and Public Health*, *9*(11), 4159–4169.

Clifton, P. M. (2011). Protein and coronary heart disease: The role of different protein sources. *Current Atherosclerosis Reports*, *13*(6), 493–498.

Colditz, G. A., Manson, J. E., Stampfer, M. J., Rosner, B., Willett, W. C., & Speizer, F. E. (1992). Diet and risk of clinical diabetes in women. *The American Journal of Clinical Nutrition*, *55*(5), 1018–1023.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Craig, W. J., & Mangels, A. R. (2009). Position of the American Dietetic Association: Vegetarian diets. *Journal of the American Dietetic Association*, *109*(7), 1266–1282.

Cross, A. J., Ferrucci, L. M., Risch, A., Graubard, B. I., Ward, M. H., Park, Y., & Sinha, R. et al. (2010). A large prospective study of meat consumption and colorectal cancer risk: An investigation of potential mechanisms underlying this association. *Cancer Research and Clinical Oncology*, *70*(6), 2406–2414.

Cross, A. J., Freedman, N. D., Ren, J., Ward, M. H., Hollenbeck, A. R., Schatzkin, A., & Abnet, C. C. et al. (2011). Meat consumption and risk of esophageal and gastric cancer in a large prospective study. *The American Journal of Gastroenterology*, *106*(3), 432–442.

Cross, A. J., & Sinha, R. (2004). Meat-related mutagens/carcinogens in the etiology of colorectal cancer. *Environmental and Molecular Mutagenesis*, *44*(1), 44–55.

Dahlquist, G. G., Blom, L. G., Persson, L. A., Sandstrom, A. I., & Wall, S. G. (1990). Dietary factors and the risk of developing insulin dependent diabetes in childhood. *British Medical Journal*, *300*(6735), 1302–1306.

Daniel, C. R., Cross, A. J., Graubard, B. I., Park, Y., Ward, M. H., Rothman, N., & Sinha, R. et al. (2012). Large prospective investigation of meat intake, related mutagens, and risk of renal cell carcinoma. *The American Journal of Clinical Nutrition*, *95*(1), 155–162.

Davis, K. E., Prasad, C., Vijayagopal, P., Juma, S., & Imrhan, V. (2014). Advanced glycation end products, inflammation, and chronic metabolic diseases: Links in a chain?[Epub ahead of print]. *Critical Reviews in Food Science and Nutrition*, (Sep): 26.

de Biase, S. G., Fernandes, S. F., Gianini, R. J., & Duarte, J. L. (2007). Vegetarian diet and cholesterol and triglycerides levels. *Arquivos Brasileiros de Cardiologia*, *88*(1), 35–39.

de Keyzer, W., van Caneghem, S., Heath, A. L., Vanaelst, B., Verschraegen, M., de Henauw, S., & Huybrechts, I. (2012). Nutritional quality and acceptability of a weekly vegetarian lunch in primary-school canteens in Ghent, Belgium: 'Thursday Veggie Day'. *Public Health Nutrition*, *15*(12), 2326–2330.

de Oliveira Otto, M. C., Alonso, A., Lee, D. H., Delclos, G. L., Bertoni, A. G., Jiang, R., & Nettleton, J. A. et al. (2012). Dietary intakes of zinc and heme iron from red meat, but not from other sources, are associated with greater risk of metabolic syndrome and cardiovascular disease. *The Journal of Nutrition*, *142*(3), 526–533.

Dellavalle, C. T., Xiao, Q., Yang, G., Shu, X. O., Aschebrook-Kilfoy, B., Zheng, W., & Ward, M. H. et al. (2014). Dietary nitrate and nitrite intake and risk of colorectal cancer in the Shanghai Women's Health Study. *International Journal of Cancer*, *134*(12), 2917–2926.

Djoussé, L., & Gaziano, J. M. (2009). Dietary cholesterol and coronary artery disease: A systematic review. *Current Atherosclerosis Reports*, *11*(6), 418–422.

Elliott, P., Stamler, J., Dyer, A. R., Appel, L., Dennis, B., Kesteloot, H., & Zhou, B. et al. (2006). Association between protein intake and blood pressure: The INTERMAP Study. *Journal of the American Medical Association Internal Medicine*, *166*(1), 79–87.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Estruch, R., Martinez-Gonzalez, M. A., Corella, D., Salas-Salvado, J., Ruiz-Gutierrez, V., Covas, M. I., & Ros, E. et al. (2006). Effects of a Mediterranean-style diet on cardiovascular risk factors: A randomized trial. *Annals of Internal Medicine*, *145*(1), 1–11.

Estruch, R., Ros, E., Salas-Salvado, J., Covas, M. I., Corella, D., Aros, F., & Martinez-Gonzalez, M. A. et al. (2013). Primary prevention of cardiovascular disease with a Mediterranean diet. *The New England Journal of Medicine*, *368*(14), 1279–1290.

Farmer, B., Larson, B. T., Fulgoni, V. L. III, Rainville, A. J., & Liepa, G. U. (2011). A vegetarian dietary pattern as a nutrient-dense approach to weight management: An analysis of the national health and nutrition examination survey 1999-2004. *Journal of the American Dietetic Association*, *111*(6), 819–827.

Farvid, M. S., Ding, M., Pan, A., Sun, Q., Chiuve, S. E., Steffen, L. M., & Hu, F. B. et al. (2014). Dietary linoleic acid and risk of coronary heart disease: A systematic review and meta-analysis of prospective cohort studies. *Circulation*, *130*(18), 1568–1578.

Fernandez-Cao, J. C., Arija, V., Aranda, N., Bullo, M., Basora, J., Martinez-Gonzalez, M. A., & Salas-Salvado, J. et al. (2013). Heme iron intake and risk of new-onset diabetes in a Mediterranean population at high risk of cardiovascular disease: An observational cohort analysis. *BioMed Central Public Health*, *13*, 1042–1048.

Ferrucci, L. M., Sinha, R., Ward, M. H., Graubard, B. I., Hollenbeck, A. R., Kilfoy, B. A., & Cross, A. J. et al. (2010). Meat and components of meat and the risk of bladder cancer in the NIH-AARP Diet and Health Study. *Cancer*, *116*(18), 4345–4353.

Feskens, E. J., & Kromhout, D. (1990). Habitual dietary intake and glucose tolerance in euglycaemic men: The Zutphen Study. *International Journal of Epidemiology*, *19*(4), 953–959.

Feskens, E. J., Sluik, D., & van Woudenbergh, G. J. (2013). Meat consumption, diabetes, and its complications. *Current Diabetes Reports*, *13*(2), 298–306.

Fonseca-Nunes, A., Jakszyn, P., & Agudo, A. (2014). Iron and cancer risk--a systematic review and meta-analysis of the epidemiological evidence. *Cancer Epidemiology, Biomarkers & Prevention*, *23*(1), 12–31.

Foo, S. Y., Heller, E. R., Wykrzykowska, J., Sullivan, C. J., Manning-Tobin, J. J., & Moore, K. J., … Rosenzweig, A. (2009). Vascular effects of a low-carbohydrate high-protein diet. *Proceedings of the National Academy of Sciences of the United States (PNAS), 106*(36), 15418-15423.

Franklin, T. L., Kolasa, K. M., Griffin, K., Mayo, C., & Badenhop, D. T. (1995). Adherance to very-low-fat diet by a group of cardiac rehabilitation patients in the rural southeastern United States. *Archives of Family Medicine*, *4*(6), 551–554.

Fraser, G. E. (1999). Associations between diet and cancer, ischemic heart disease, and all-cause mortality in non-Hispanic white California Seventh-day Adventists. *The American Journal of Clinical Nutrition*, *70*(3), 532S–538S.

Fraser, G. E. (2009). Vegetarian diets: What do we know of their effects on common chronic diseases? *The American Journal of Clinical Nutrition*, *89*(5), 1607S–1612S.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Fraser, G. E., & Shavlik, D. J. (2001). Ten years of life: Is it a matter of choice? *Journal of the American Medical Association Internal Medicine*, *161*(13), 1645–1652.

Friel, S., Dangour, A. D., Garnett, T., Lock, K., Chalabi, Z., Roberts, I., & Haines, A. et al. (2009). Public health benefits of strategies to reduce greenhouse-gas emissions: Food and agriculture. *Lancet*, *374*(9706), 2016–2025.

Fung, T. T., Hu, F. B., Hankinson, S. E., Willett, W. C., & Holmes, M. D. (2011). Low-carbohydrate diets, dietary approaches to stop hypertension-style diets, and the risk of postmenopausal breast cancer. *American Journal of Epidemiology*, *174*(6), 652–660.

Fung, T. T., Hu, F. B., Wu, K., Chiuve, S. E., Fuchs, C. S., & Giovannucci, E. (2010). The Mediterranean and Dietary Approaches to Stop Hypertension (DASH) diets and colorectal cancer. *The American Journal of Clinical Nutrition*, *92*(6), 1429–1435.

Fung, T. T., McCullough, M., van Dam, R. M., & Hu, F. B. (2007). A prospective study of overall diet quality and risk of type 2 diabetes in women. *Diabetes Care*, *30*(7), 1753–1757.

Fung, T. T., McCullough, M. L., Newby, P. K., Manson, J. E., Meigs, J. B., Rifai, N., & Hu, F. B. et al. (2005). Diet-quality scores and plasma concentrations of markers of inflammation and endothelial dysfunction. *The American Journal of Clinical Nutrition*, *82*(1), 163–173.

Fung, T. T., van Dam, R. M., Hankinson, S. E., Stampfer, M., Willett, W. C., & Hu, F. B. (2010). Low-carbohydrate diets and all-cause and cause-specific mortality: Two cohort studies. *Annals of Internal Medicine*, *153*(5), 289–298.

García-Fernández, E., Rico-Cabanas, L., Rosgaard, N., Estruch, R., & Bach-Faig, A. (2014). Mediterranean diet and cardiodiabesity: A review. *Nutrients*, *6*(9), 3474–3500.

Gigleux, I., Jenkins, D. J., Kendall, C. W., Marchie, A., Faulkner, D. A., Wong, J. M., & Lamarche, B. et al. (2007). Comparison of a dietary portfolio diet of cholesterol-lowering foods and a statin on LDL particle size phenotype in hypercholesterolaemic participants. *The British Journal of Nutrition*, *98*(6), 1229–1236.

Gilsing, A. M., Fransen, F., de Kok, T. M., Goldbohm, A. R., Schouten, L. J., de Bruine, A. P., & Weijenberg, M. P. et al. (2013). Dietary heme iron and the risk of colorectal cancer with specific mutations in KRAS and APC. *Carcinogenesis*, *34*(12), 2757–2766.

Glick-Bauer, M., & Yeh, M. C. (2014). The Health Advantage of a Vegan Diet: Exploring the Gut Microbiota Connection. *Nutrients*, *6*(11), 4822–4838.

Habermeyer, M., Roth, A., Guth, S., Diel, P., Engel, K. H., Epe, B., & Eisenbrand, G. et al. (2014). Nitrate and nitrite in the diet: How to assess their benefit and risk for human health. *Molecular Nutrition & Food Research*, *59*(1), 106–128.

Haddad, E. H., & Tanzman, J. S. (2003). What do vegetarians in the United States eat? *The American Journal of Clinical Nutrition*, *78*(3), 626S–632S.

Halkjaer, J., Olsen, A., Overvad, K., Jakobsen, M. U., Boeing, H., Buijsse, B., & Tjonneland, A. et al. (2011). Intake of total, animal and plant protein and subsequent changes in weight or waist circumference in European men and women: The Diogenes project. *International Journal of Obesity*, *35*(8), 1104–1113.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Helgason, T., Ewen, S. W., Ross, I. S., & Stowers, J. M. (1982). Diabetes produced in mice by smoked/cured mutton. *Lancet*, *320*(8306), 1017–1022.

Holtrop, G., Johnstone, A. M., Fyfe, C., & Gratz, S. W. (2012). Diet composition is associated with endogenous formation of N-nitroso compounds in obese men. *The Journal of Nutrition*, *142*(9), 1652–1658.

Hooda, J., Shah, A., & Zhang, L. (2014). Heme, an essential nutrient from dietary proteins, critically impacts diverse physiological and pathological processes. *Nutrients*, *6*(3), 1080–1102.

Hua, N. W., Stoohs, R. A., & Facchini, F. S. (2001). Low iron status and enhanced insulin sensitivity in lacto-ovo vegetarians. *The British Journal of Nutrition*, *86*(4), 515–519.

Huang, T., Yang, B., Zheng, J., Li, G., Wahlqvist, M. L., & Li, D. (2012). Cardiovascular disease mortality and cancer incidence in vegetarians: A meta-analysis and systematic review. *Annals of Nutrition & Metabolism*, *60*(4), 233–240.

Hughes, R., Cross, A. J., Pollock, J. R., & Bingham, S. (2001). Dose-dependent effect of dietary meat on endogenous colonic N-nitrosation. *Carcinogenesis*, *22*(1), 199–202.

Hunnicutt, J., He, K., & Xun, P. (2014). Dietary iron intake and body iron stores are associated with risk of coronary heart disease in a meta-analysis of prospective cohort studies. *The Journal of Nutrition*, *144*(3), 359–366.

Jakobsen, M. U., Dethlefsen, C., Joensen, A. M., Stegger, J., Tjønneland, A., Schmidt, E. B., & Overvad, K. (2010). Intake of carbohydrates compared with intake of saturated fatty acids and risk of myocardial infarction: Importance of the glycemic index. *The American Journal of Clinical Nutrition*, *91*(6), 1764–1768.

Jakobsen, M. U., O'Reilly, E. J., Heitmann, B. L., Pereira, M. A., Bälter, K., Fraser, G. E., & Ascherio, A. et al. (2009). Major types of dietary fat and risk of coronary heart disease: A pooled analysis of 11 cohort studies. *The American Journal of Clinical Nutrition*, *89*(5), 1425–1432.

Jakobsen, M. U., Overvad, K., Dyerberg, J., & Heitmann, B. L. (2008). Intake of ruminant trans fatty acids and risk of coronary heart disease. *International Journal of Epidemiology*, *37*(1), 173–182.

Jamin, E. L., Riu, A., Douki, T., Debrauwer, L., Cravedi, J. P., Zalko, D., & Audebert, M. (2013). Combined genotoxic effects of a polycyclic aromatic hydrocarbon (B(a)P) and an heterocyclic amine (PhIP) in relation to colorectal carcinogenesis. *Public Library of Science (PLoS). One*, *8*(3), 1–11.

Jenkins, D. J., Kendall, C. W., Faulkner, D., Vidgen, E., Trautwein, E. A., Parker, T. L., & Connelly, P. W. et al. (2002). A dietary portfolio approach to cholesterol reduction: Combined effects of plant sterols, vegetable proteins, and viscous fibers in hypercholesterolemia. *Metabolism: Clinical and Experimental*, *51*(12), 1596–1604.

Jenkins, D. J., Kendall, C. W., Faulkner, D. A., Kemp, T., Marchie, A., Nguyen, T. H., & Singer, W. et al. (2008). Long-term effects of a plant-based dietary portfolio of cholesterol-lowering foods on blood pressure. *European Journal of Clinical Nutrition*, *62*(6), 781–788.

Jenkins, D. J., Kendall, C. W., Marchie, A., Faulkner, D. A., Josse, A. R., Wong, J. M., & Connelly, P. W. et al. (2005). Direct comparison of dietary portfolio vs statin on C-reactive protein. *European Journal of Clinical Nutrition*, *59*(7), 851–860.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Jenkins, D. J., Kendall, C. W., Marchie, A., Faulkner, D. A., Wong, J. M., de Souza, R., & Connelly, P. W. et al. (2003). Effects of a dietary portfolio of cholesterol-lowering foods vs lovastatin on serum lipids and C-reactive protein. *Journal of the American Medical Association*, *290*(4), 502–510.

Jenkins, D. J., Kendall, C. W., Marchie, A., Jenkins, A. L., Augustin, L. S., Ludwig, D. S., & Anderson, J. W. et al. (2003). Type 2 diabetes and the vegetarian diet. *The American Journal of Clinical Nutrition*, *78*(3), 610S–616S.

Jenkins, D. J., Kendall, C. W., Nguyen, T. H., Marchie, A., Faulkner, D. A., Ireland, C., & Singer, W. et al. (2008). Effect of plant sterols in combination with other cholesterol-lowering foods. *Metabolism: Clinical and Experimental*, *57*(1), 130–139.

Jiang, R., Manson, J. E., Meigs, J. B., Ma, J., Rifai, N., & Hu, F. B. (2004). Body iron stores in relation to risk of type 2 diabetes in apparently healthy women. *Journal of the American Medical Association*, *291*(6), 711–717.

Jiao, L., Kramer, J. R., Chen, L., Rugge, M., Parente, P., Verstovsek, G., & El-Serag, H. B. et al. (2013). Dietary consumption of meat, fat, animal products and advanced glycation end-products and the risk of Barrett's oesophagus. *Alimentary Pharmacology & Therapeutics*, *38*(7), 817–824.

Jiao, L., Stolzenberg-Solomon, R., Zimmerman, T. P., Duan, Z., Chen, L., Kahle, L., & Sinha, R. et al. (2015). Dietary consumption of advanced glycation end products and pancreatic cancer in the prospective NIH-AARP Diet and Health Study. *The American Journal of Clinical Nutrition*, *101*(1), 126–134.

Johnson, C. M., Wei, C., Ensor, J. E., Smolenski, D. J., Amos, C. I., Levin, B., & Berry, D. A. (2013). Meta-analyses of colorectal cancer risk factors. *Cancer Causes & Control*, *24*(6), 1207–1222.

Kahleova, H., Matoulek, M., Malinska, H., Oliyarnik, O., Kazdova, L., Neskudla, T., & Pelikanova, T. et al. (2011). Vegetarian diet improves insulin resistance and oxidative stress markers more than conventional diet in subjects with Type 2 diabetes. *Diabetic Medicine*, *28*(5), 549–559.

Kaluza, J., Akesson, A., & Wolk, A. (2014). Processed and unprocessed red meat consumption and risk of heart failure: Prospective study of men. *Circulation: Heart Failure*, *7*(4), 552–557.

Kaluza, J., Larsson, S. C., Hakansson, N., & Wolk, A. (2014). Heme iron intake and acute myocardial infarction: A prospective study of men. *International Journal of Cardiology*, *172*(1), 155–160.

Kaluza, J., Wolk, A., & Larsson, S. C. (2012). Red meat consumption and risk of stroke: A meta-analysis of prospective studies. *Stroke*, *43*(10), 2556–2560.

Kaluza, J., Wolk, A., & Larsson, S. C. (2013). Heme iron intake and risk of stroke: A prospective study of men. *Stroke*, *44*(2), 334–339.

Kastorini, C. M., Milionis, H. J., Esposito, K., Giugliano, D., Goudevenos, J. A., & Panagiotakos, D. B. (2011). The effect of Mediterranean diet on metabolic syndrome and its components: A meta-analysis of 50 studies and 534,906 individuals. *Journal of the American College of Cardiology*, *57*(11), 1299–1313.

Katcher, H. I., Ferdowsian, H. R., Hoover, V. J., Cohen, J. L., & Barnard, N. D. (2010). A worksite vegan nutrition program is well-accepted and improves health-related quality of life and work productivity. *Annals of Nutrition & Metabolism*, *56*(4), 245–252.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Kellow, N. J., & Savige, G. S. (2013). Dietary advanced glycation end-product restriction for the attenuation of insulin resistance, oxidative stress and endothelial dysfunction: A systematic review. *European Journal of Clinical Nutrition*, *67*(3), 239–248.

Keszei, A. P., Goldbohm, R. A., Schouten, L. J., Jakszyn, P., & van den Brandt, P. A. (2013). Dietary N-nitroso compounds, endogenous nitrosation, and the risk of esophageal and gastric cancer subtypes in the Netherlands Cohort Study. *The American Journal of Clinical Nutrition*, *97*(1), 135–146.

Key, T. J., Appleby, P. N., Spencer, E. A., Travis, R. C., Roddam, A. W., & Allen, N. E. (2009). Mortality in British vegetarians: Results from the European Prospective Investigation into Cancer and Nutrition (EPIC-Oxford). *The American Journal of Clinical Nutrition*, *89*(5), 1613S–1619S.

Key, T. J., Fraser, G. E., Thorogood, M., Appleby, P. N., Beral, V., Reeves, G., & McPherson, K. et al. (1998). Mortality in vegetarians and non-vegetarians: A collaborative analysis of 8300 deaths among 76,000 men and women in five prospective studies. *Public Health Nutrition*, *1*(1), 33–41.

Key, T. J., Fraser, G. E., Thorogood, M., Appleby, P. N., Beral, V., Reeves, G., & McPherson, K. et al. (1999). Mortality in vegetarians and nonvegetarians: Detailed findings from a collaborative analysis of 5 prospective studies. *The American Journal of Clinical Nutrition*, *70*(3Suppl), 516S–524S.

Kim, E., Coelho, D., & Blachier, F. (2013). Review of the association between meat consumption and risk of colorectal cancer. *Nutrition Research (New York, N.Y.)*, *33*(12), 983–994.

Knowler, W. C., Barrett-Connor, E., Fowler, S. E., Hamman, R. F., Lachin, J. M., Walker, E. A., & Nathan, D. M. (2002). Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *The New England Journal of Medicine*, *346*(6), 393–403.

Koeth, R. A., Wang, Z., Levison, B. S., Buffa, J. A., Org, E., Sheehy, B. T., & Hazen, S. L. et al. (2013). Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nature Medicine*, *19*(5), 576–585.

Kontogianni, M. D., Panagiotakos, D. B., Pitsavos, C., Chrysohoou, C., & Stefanadis, C. (2008). Relationship between meat intake and the development of acute coronary syndromes: The CARDIO2000 case-control study. *European Journal of Clinical Nutrition*, *62*(2), 171–177.

Kratz, M. (2005). Dietary cholesterol, atherosclerosis and coronary heart disease. *Handbook of Experimental Pharmacology*, *170*, 195–213.

Kromhout, D., Geleijnse, J. M., Menotti, A., & Jacobs, D. R. Jr. (2011). The confusion about dietary fatty acids recommendations for CHD prevention. *The British Journal of Nutrition*, *106*(5), 627–632.

Kunutsor, S. K., Apekey, T. A., Walley, J., & Kain, K. (2013). Ferritin levels and risk of type 2 diabetes mellitus: An updated systematic review and meta-analysis of prospective evidence. *Diabetes/Metabolism Research and Reviews*, *29*(4), 308–318.

Larsson, S. C., & Orsini, N. (2014). Red meat and processed meat consumption and all-cause mortality: A meta-analysis. *American Journal of Epidemiology*, *179*(3), 282–289.

Larsson, S. C., & Wolk, A. (2012). Red and processed meat consumption and risk of pancreatic cancer: Meta-analysis of prospective studies. *British Journal of Cancer*, *106*(3), 603–607.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Lasa, A., Miranda, J., Bullo, M., Casas, R., Salas-Salvado, J., Larretxi, I., & Portillo, M. P. et al. (2014). Comparative effect of two Mediterranean diets versus a low-fat diet on glycaemic control in individuals with type 2 diabetes. *European Journal of Clinical Nutrition*, *68*(7), 767–772.

Le, L. T., & Sabate, J. (2014). Beyond meatless, the health effects of vegan diets: Findings from the Adventist cohorts. *Nutrients*, *6*(6), 2131–2147.

Lee, D. H., Folsom, A. R., & Jacobs, D. R. (2004). Dietary iron intake and Type 2 diabetes incidence in postmenopausal women: The Iowa Women s Health Study. *Diabetologia*, *47*(2), 185–194.

Lee, J. E., Spiegelman, D., Hunter, D. J., Albanes, D., Bernstein, L., van den Brandt, P. A., & Smith-Warner, S. A. et al. (2008). Fat, protein, and meat consumption and renal cell cancer risk: A pooled analysis of 13 prospective studies. *Journal of the National Cancer Institute*, *100*(23), 1695–1706.

Lenzen, S. (2008). Oxidative stress: The vulnerable beta-cell. *Biochemical Society Transactions*, *36*(3), 343–347.

Lichtenstein, A. H., Appel, L. J., Brands, M., Carnethon, M., Daniels, S., Franch, H. A., & Wylie-Rosett, J. et al. (2006). Summary of American Heart Association Diet and Lifestyle Recommendations revision 2006. *Arteriosclerosis, Thrombosis, and Vascular Biology*, *26*(10), 2186–2191.

Lindström, J., Peltonen, M., Eriksson, J. G., Louheranta, A., Fogelholm, M., Uusitupa, M., & Tuomilehto, J. (2006). High-fibre, low-fat diet predicts long-term weight loss and decreased type 2 diabetes risk: The Finnish Diabetes Prevention Study. *Diabetologia*, *49*(5), 912–920.

Lovejoy, J., & DiGirolamo, M. (1992). Habitual dietary intake and insulin sensitivity in lean and obese adults. *The American Journal of Clinical Nutrition*, *55*(6), 1174–1179.

Lu, S. C., Wu, W. H., Lee, C. A., Chou, H. F., Lee, H. R., & Huang, P. C. (2000). LDL of Taiwanese vegetarians are less oxidizable than those of omnivores. *The Journal of Nutrition*, *130*(6), 1591–1596.

Luan de, C., Li, H., Li, S. J., Zhao, Z., Li, X., & Liu, Z. M. (2008). Body iron stores and dietary iron intake in relation to diabetes in adults in North China. *Diabetes Care*, *31*(2), 285–286.

Marsh, K., & Brand-Miller, J. (2011). Vegetarian Diets and Diabetes. *American Journal of Lifestyle Medicine*, *5*(2), 135–143.

Marshall, J. A., Bessesen, D. H., & Hamman, R. F. (1997). High saturated fat and low starch and fibre are associated with hyperinsulinaemia in a non-diabetic population: The San Luis Valley Diabetes Study. *Diabetologia*, *40*(4), 430–438.

Marshall, J. A., Hamman, R. F., & Baxter, J. (1991). High-fat, low-carbohydrate diet and the etiology of non-insulin-dependent diabetes mellitus: The San Luis Valley Diabetes Study. *American Journal of Epidemiology*, *134*(6), 590–603.

National Cancer Institute (NCI). (2013). Sources of saturated fat in the diet of the U.S. population ages 2 years and older, NHANES 2005-2006. Retrieved from http://appliedresearch.cancer.gov/diet/food-sources/sat_fat.html

Marshall, J. A., Hoag, S., Shetterly, S., & Hamman, R. F. (1994). Dietary fat predicts conversion from impaired glucose tolerance to NIDDM. The San Luis Valley Diabetes Study. *Diabetes Care*, *17*(1), 50–56.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Martinez-Gonzalez, M. A., & Bes-Rastrollo, M. (2014). Dietary patterns, Mediterranean diet, and cardiovascular disease. *Current Opinion in Lipidology*, *25*(1), 20–26.

Mayer, E. J., Newman, B., Quesenberry, C. P., & Selby, J. V. (1993). Usual dietary fat intake and insulin concentrations in healthy women twins. *Diabetes Care*, *16*(11), 1459–1469.

Mayer-Davis, E. J., Monaco, J. H., Hoen, H. M., Carmichael, S., Vitolins, M. Z., Rewers, M. J., & Karter, A. J. et al. (1997). Dietary fat and insulin sensitivity in a triethnic population: The role of obesity. The Insulin Resistance Atherosclerosis Study (IRAS). *The American Journal of Clinical Nutrition*, *65*(1), 79–87.

McCullough, M. L., Feskanich, D., Stampfer, M. J., Giovannucci, E. L., Rimm, E. B., Hu, F. B., & Willett, W. C. et al. (2002). Diet quality and major chronic disease risk in men and women: Moving toward improved dietary guidance. *The American Journal of Clinical Nutrition*, *76*(6), 1261–1271.

McEvoy, C. T., Temple, N., & Woodside, J. V. (2012). Vegetarian diets, low-meat diets and health: A review. *Public Health Nutrition*, *15*(12), 2287–2294.

Mensink, R. P., Zock, P. L., Katan, M. B., & Hornstra, G. (1992). Effect of dietary cis and trans fatty acids on serum lipoprotein[a] levels in humans. *Journal of Lipid Research*, *33*(10), 1493–1501.

Micha, R., Michas, G., Lajous, M., & Mozaffarian, D. (2013). Processing of meats and cardiovascular risk: Time to focus on preservatives. *BioMed Central Medicine*, *11*, 136. doi:10.1186/1741-7015-11-136

Micha, R., Michas, G., & Mozaffarian, D. (2012). Unprocessed red and processed meats and risk of coronary artery disease and type 2 diabetes--an updated review of the evidence. *Current Atherosclerosis Reports*, *14*(6), 515–524.

Micha, R., Wallace, S. K., & Mozaffarian, D. (2010). Red and processed meat consumption and risk of incident coronary heart disease, stroke, and diabetes mellitus: A systematic review and meta-analysis. *Circulation*, *121*(21), 2271–2283.

Mozaffarian, D., Aro, A., & Willett, W. C. (2009). Health effects of trans-fatty acids: Experimental and observational evidence. *European Journal of Clinical Nutrition*, *63*(2), S5–S21.

Mozaffarian, D., Hao, T., Rimm, E. B., Willett, W. C., & Hu, F. B. (2011). Changes in diet and lifestyle and long-term weight gain in women and men. *The New England Journal of Medicine*, *364*(25), 2392–2404.

Mozaffarian, D., Katan, M. B., Ascherio, A., Stampfer, M. J., & Willett, W. C. (2006). Trans fatty acids and cardiovascular disease. *The New England Journal of Medicine*, *354*(15), 1601–1613.

Mozaffarian, D., Micha, R., & Wallace, S. (2010). Effects on coronary heart disease of increasing polyunsaturated fat in place of saturated fat: A systematic review and meta-analysis of randomized controlled trials. *Public Library of Science (PLoS). Medicine*, *7*(3), 1–10.

Mozaffarian, D., Pischon, T., Hankinson, S. E., Rifai, N., Joshipura, K., Willett, W. C., & Rimm, E. B. (2004). Dietary intake of trans fatty acids and systemic inflammation in women. *The American Journal of Clinical Nutrition*, *79*(4), 606–612.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

National Cancer Institute (NCI). (2013). Sources of saturated fat in the diets of the U.S. population ages 2 years and older, NHANES 2005-2006. Retrieved November 17, 2014, from http://appliedresearch.cancer.gov/diet/foodsources/sat_fat/sf.html

National Health and Medical Research Council (NHMRC). (2013). Australian Dietary Guidelines. Canberra: Australian Government, NHMRC, Department of Health and Ageing. Retrieved September 8, 2015, from https://www.nhmrc.gov.au/guidelines-publications/n55

National Institute of Health (NIH). (2014). Ferritin blood tests. Retrieved 17 November, 2014, from http://www.nlm.nih.gov/medlineplus/ency/article/003490.htm

Nicholson, A. S., Sklar, M., Barnard, N. D., Gore, S., Sullivan, R., & Browning, S. (1999). Toward improved management of NIDDM: A randomized, controlled, pilot intervention using a lowfat, vegetarian diet. *Preventive Medicine*, *29*(2), 87–91.

Nieman, D. C., Sherman, K. M., Arabatzis, K., Underwood, B. C., Barbosa, J. C., Johnson, M., & Lee, J. et al. (1989). Hematological, anthropometric, and metabolic comparisons between vegetarian and nonvegetarian elderly women. *International Journal of Sports Medicine*, *10*(4), 243–251.

Noto, H., Goto, A., Tsujimoto, T., & Noda, M. (2013). Low-carbohydrate diets and all-cause mortality: A systematic review and meta-analysis of observational studies. *Public Library of Science (PLoS). One*, *8*(1), 1–10.

Oomen, C. M., Ocke, M. C., Feskens, E. J., van Erp-Baart, M. A., Kok, F. J., & Kromhout, D. (2001). Association between trans fatty acid intake and 10-year risk of coronary heart disease in the Zutphen Elderly Study: A prospective population-based study. *Lancet*, *357*(9258), 746–751.

Opara, E. C. (2004). Role of oxidative stress in the etiology of type 2 diabetes and the effect of antioxidant supplementation on glycemic control. *Journal of Investigative Medicine*, *52*(1), 19–23.

Orlich, M. J., Singh, P. N., Sabate, J., Jaceldo-Siegl, K., Fan, J., Knutsen, S., & Fraser, G. E. et al. (2013). Vegetarian dietary patterns and mortality in Adventist Health Study 2. *Journal of the American Medical Association Internal Medicine*, *173*(13), 1230–1238.

Padler-Karavani, V., Yu, H., Cao, H., Chokhawala, H., Karp, F., Varki, N., & Varki, A. et al. (2008). Diversity in specificity, abundance, and composition of anti-Neu5Gc antibodies in normal humans: Potential implications for disease. *Glycobiology*, *18*(10), 818–830.

Paluszkiewicz, P., Smolinska, K., Debinska, I., & Turski, W. A. (2012). Main dietary compounds and pancreatic cancer risk. The quantitative analysis of case-control and cohort studies. *Cancer Epidemiology*, *36*(1), 60–67.

Pan, A., Sun, Q., Bernstein, A. M., Manson, J. E., Willett, W. C., & Hu, F. B. (2013). Changes in red meat consumption and subsequent risk of type 2 diabetes mellitus: Three cohorts of US men and women. *Journal of the American Medical Association Internal Medicine*, *173*(14), 1328–1335.

Pan, A., Sun, Q., Bernstein, A. M., Schulze, M. B., Manson, J. E., Stampfer, M. J., & Hu, F. B. et al. (2012). Red meat consumption and mortality: Results from 2 prospective cohort studies.[formerly Archives of Internal Medicine]. *Journal of the American Medical Association Internal Medicine*, *172*(7), 555–563.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Pan, A., Sun, Q., Bernstein, A. M., Schulze, M. B., Manson, J. E., Willett, W. C., & Hu, F. B. (2011). Red meat consumption and risk of type 2 diabetes: 3 cohorts of US adults and an updated meta-analysis. *The American Journal of Clinical Nutrition*, *94*(4), 1088–1096.

Parillo, M., & Riccardi, G. (2004). Diet composition and the risk of type 2 diabetes: Epidemiological and clinical evidence. *The British Journal of Nutrition*, *92*(1), 7–19.

Parker, D. R., Weiss, S. T., Troisi, R., Cassano, P. A., Vokonas, P. S., & Landsberg, L. (1993). Relationship of dietary saturated fatty acids and body habitus to serum insulin concentrations: The Normative Aging Study. *The American Journal of Clinical Nutrition*, *58*(2), 129–136.

Pereira, M. A., Jacobs, D. R. Jr, Van Horn, L., Slattery, M. L., Kartashov, A. I., & Ludwig, D. S. (2002). Dairy consumption, obesity, and the insulin resistance syndrome in young adults: The CARDIA Study. *Journal of the American Medical Association*, *287*(16), 2081–2089.

Pham, N. M., Mizoue, T., Tanaka, K., Tsuji, I., Tamakoshi, A., Matsuo, K., & Sasazuki, S. et al. (2014). Meat consumption and colorectal cancer risk: An evaluation based on a systematic review of epidemiologic evidence among the Japanese population. *Japanese Journal of Clinical Oncology*, *44*(7), 641–650.

Pietinen, P., Ascherio, A., Korhonen, P., Hartman, A. M., Willett, W. C., Albanes, D., & Virtamo, J. (1997). Intake of fatty acids and risk of coronary heart disease in a cohort of Finnish men. The Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study. *American Journal of Epidemiology*, *145*(10), 876–887.

Portha, B., Giroix, M. H., Cros, J. C., & Picon, L. (1980). Diabetogenic effect of N-nitrosomethylurea and N-nitrosomethylurethane in the adult rat. *Annales de la Nutrition et de l'Alimentation*, *34*(5-6), 1143–1151.

Pounis, G. D., Tyrovolas, S., Antonopoulou, M., Zeimbekis, A., Anastasiou, F., Bountztiouka, V., & Panagiotakos, D. B. et al. (2010). Long-term animal-protein consumption is associated with an increased prevalence of diabetes among the elderly: The Mediterranean islands (MEDIS) study. *Diabetes & Metabolism*, *36*(6), 484–490.

Pratt, M. M., John, K., MacLean, A. B., Afework, S., Phillips, D. H., & Poirier, M. C. (2011). Polycyclic aromatic hydrocarbon (PAH) exposure and DNA adduct semi-quantitation in archived human tissues. *International Journal of Environmental Research and Public Health*, *8*(7), 2675–2691.

Preis, S. R., Stampfer, M. J., Spiegelman, D., Willett, W. C., & Rimm, E. B. (2010). Dietary protein and risk of ischemic heart disease in middle-aged men. *The American Journal of Clinical Nutrition*, *92*(5), 1265–1272.

Qi, L., van Dam, R. M., Rexrode, K. M., & Hu, F. B. (2007). Heme iron from diet as a risk factor for coronary heart disease in women with type 2 diabetes. *Diabetes Care*, *30*(1), 101–106.

Qiu, C., Zhang, C., Gelaye, B., Enquobahrie, D. A., Frederick, I. O., & Williams, M. A. (2011). Gestational diabetes mellitus in relation to maternal dietary heme iron and nonheme iron intake. *Diabetes Care*, *34*(7), 1564–1569.

Rajpathak, S., Ma, J., Manson, J., Willett, W. C., & Hu, F. B. (2006). Iron intake and the risk of type 2 diabetes in women: A prospective cohort study. *Diabetes Care*, *29*(6), 1370–1376.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Rajpathak, S. N., Crandall, J. P., Wylie-Rosett, J., Kabat, G. C., Rohan, T. E., & Hu, F. B. (2009). The role of iron in type 2 diabetes in humans. *Biochimica et Biophysica Acta*, *1790*(7), 671–681.

Reid, M. A., Marsh, K. A., Zeuschner, C. L., Saunders, A. V., & Baines, S. K. (2013). Meeting the nutrient reference values on a vegetarian diet. *The Medical Journal of Australia*, *199*(4), S33–S40.

Rizzo, N. S., Jaceldo-Siegl, K., Sabate, J., & Fraser, G. E. (2013). Nutrient profiles of vegetarian and nonvegetarian dietary patterns. *Journal of the Academy of Nutrition and Dietetics*, *113*(12), 1610–1619.

Rizzo, N. S., Sabate, J., Jaceldo-Siegl, K., & Fraser, G. E. (2011). Vegetarian dietary patterns are associated with a lower risk of metabolic syndrome: The adventist health study 2. *Diabetes Care*, *34*(5), 1225–1227.

Rohrmann, S., Overvad, K., Bueno-de-Mesquita, H. B., Jakobsen, M. U., Egeberg, R., Tjonneland, A., & Linseisen, J. et al. (2013). Meat consumption and mortality--results from the European Prospective Investigation into Cancer and Nutrition. *BioMed Central Medicine*, *11*(63), 1–12.

Romaguera, D., Norat, T., Vergnaud, A. C., Mouw, T., May, A. M., Agudo, A., & Peeters, P. H. et al. (2010). Mediterranean dietary patterns and prospective weight change in participants of the EPIC-PANACEA project. *The American Journal of Clinical Nutrition*, *92*(4), 912–921.

Rosell, M., Appleby, P., Spencer, E., & Key, T. (2006). Weight gain over 5 years in 21,966 meat-eating, fish-eating, vegetarian, and vegan men and women in EPIC-Oxford. *International Journal of Obesity*, *30*(9), 1389–1396.

Rouse, I. L., Armstrong, B. K., & Beilin, L. J. (1983). The relationship of blood pressure to diet and lifestyle in two religious populations. *Journal of Hypertension*, *1*(1), 65–71.

Russell, W. R., Gratz, S. W., Duncan, S. H., Holtrop, G., Ince, J., Scobbie, L., & Flint, H. J. et al. (2011). High-protein, reduced-carbohydrate weight-loss diets promote metabolite profiles likely to be detrimental to colonic health. *The American Journal of Clinical Nutrition*, *93*(5), 1062–1072.

Salas-Salvadó, J., Bulló, M., Babio, N., Martínez-González, M. A., Ibarrola-Jurado, N., Basora, J., & Ros, E. et al. (2011). Reduction in the Incidence of Type 2 Diabetes with the Mediterranean Diet. *Diabetes Care*, *34*(1), 14–19.

Salehi, M., Moradi-Lakeh, M., Salehi, M. H., Nojomi, M., & Kolahdooz, F. (2013). Meat, fish, and esophageal cancer risk: A systematic review and dose-response meta-analysis. *Nutrition Reviews*, *71*(5), 257–267.

Salehi-Abargouei, A., Maghsoudi, Z., Shirani, F., & Azadbakht, L. (2013). Effects of Dietary Approaches to Stop Hypertension (DASH)-style diet on fatal or nonfatal cardiovascular diseases--incidence: A systematic review and meta-analysis on observational prospective studies. *Nutrition (Burbank, Los Angeles County, Calif.)*, *29*(4), 611–618.

Samraj, A. N., Laubli, H., Varki, N., & Varki, A. (2014). Involvement of a non-human sialic acid in human cancer. *Frontiers in Oncology*, *4*, 1–13.

Saunders, A. V., Craig, W. J., Baines, S. K., & Posen, J. S. (2012). Iron and vegetarian diets. *The Medical Journal of Australia*, *1*(2), 11–16.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Scarborough, P., Allender, S., Clarke, D., Wickramasinghe, K., & Rayner, M. (2012). Modelling the health impact of environmentally sustainable dietary scenarios in the UK. *European Journal of Clinical Nutrition*, *66*(6), 710–715.

Schulz, M., Kroke, A., Liese, A. D., Hoffmann, K., Bergmann, M. M., & Boeing, H. (2002). Food groups as predictors for short-term weight changes in men and women of the EPIC-Potsdam cohort. *The Journal of Nutrition*, *132*(6), 1335–1340.

Schwab, U., Lauritzen, L., Tholstrup, T., Haldorssoni, T., Riserus, U., Uusitupa, M., & Becker, W. (2014). Effect of the amount and type of dietary fat on cardiometabolic risk factors and risk of developing type 2 diabetes, cardiovascular diseases, and cancer: A systematic review. *Food & Nutrition Research*, *58*, 1–26.

Schwab, U. S., Ausman, L. M., Vogel, S., Li, Z., Lammi-Keefe, C. J., Goldin, B. R., & Lichtenstein, A. H. et al. (2000). Dietary cholesterol increases the susceptibility of low density lipoprotein to oxidative modification. *Atherosclerosis*, *149*(1), 83–90.

Shirani, F., Salehi-Abargouei, A., & Azadbakht, L. (2013). Effects of Dietary Approaches to Stop Hypertension (DASH) diet on some risk for developing type 2 diabetes: A systematic review and meta-analysis on controlled clinical trials. *Nutrition (Burbank, Los Angeles County, Calif.)*, *29*(7-8), 939–947.

Sievenpiper, J. L., & Dworatzek, P. D. (2013). Food and dietary pattern-based recommendations: An emerging approach to clinical practice guidelines for nutrition therapy in diabetes. *Canadian Journal of Diabetes*, *37*(1), 51–57.

Singh, P. N., Sabate, J., & Fraser, G. E. (2003). Does low meat consumption increase life expectancy in humans? *The American Journal of Clinical Nutrition*, *78*(3), 526S–532S.

Sinha, R., Cross, A. J., Graubard, B. I., Leitzmann, M. F., & Schatzkin, A. (2009). Meat intake and mortality: A prospective study of over half a million people. *Journal of the American Medical Association Internal Medicine*, *169*(6), 562–571.

Sinha, R., Peters, U., Cross, A. J., Kulldorff, M., Weissfeld, J. L., Pinsky, P. F., & Hayes, R. B. et al. (2005). Meat, meat cooking methods and preservation, and risk for colorectal adenoma. *Cancer Research*, *65*(17), 8034–8041.

Sipetic, S. B., Vlajinac, H. D., Kocev, N. I., Marinkovic, J. M., Radmanovic, S. Z., & Bjekic, M. D. (2005). The Belgrade childhood diabetes study: A multivariate analysis of risk determinants for diabetes. *European Journal of Public Health*, *15*(2), 117–122.

Sluijs, I., & Beulens, J. W., van der A, D. L., Spijkerman, A. M., Grobbee, D. E., & van der Schouw, Y. T. (2010). Dietary intake of total, animal, and vegetable protein and risk of type 2 diabetes in the European Prospective Investigation into Cancer and Nutrition (EPIC)-NL study. *Diabetes Care*, *33*(1), 43–48.

Smolinska, K., & Paluszkiewicz, P. (2010). Risk of colorectal cancer in relation to frequency and total amount of red meat consumption. Systematic review and meta-analysis. *Archives of Medical Science*, *6*(4), 605–610.

Snowdon, D. A., & Phillips, R. L. (1985). Does a vegetarian diet reduce the occurrence of diabetes? *American Journal of Public Health*, *75*(5), 507–512.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Snowdon, D. A., Phillips, R. L., & Fraser, G. E. (1984). Meat consumption and fatal ischemic heart disease. *Preventive Medicine*, *13*(5), 490–500.

Sofi, F., Macchi, C., Abbate, R., Gensini, G. F., & Casini, A. (2014). Mediterranean diet and health status: An updated meta-analysis and a proposal for a literature-based adherence score. *Public Health Nutrition*, *17*(12), 2769–2782.

Song, P., Lu, M., Yin, Q., Wu, L., Zhang, D., Fu, B., & Zhao, Q. et al. (2014). Red meat consumption and stomach cancer risk: A meta-analysis. *Journal of Cancer Research and Clinical Oncology*, *140*(6), 979–992.

Spencer, E. A., Appleby, P. N., Davey, G. K., & Key, T. J. (2003). Diet and body mass index in 38000 EPIC-Oxford meat-eaters, fish-eaters, vegetarians and vegans. *International Journal of Obesity and Related Metabolic Disorders*, *27*(6), 728–734.

Tangvoranuntakul, P., Gagneux, P., Diaz, S., Bardor, M., Varki, N., Varki, A., & Muchmore, E. (2003). Human uptake and incorporation of an immunogenic nonhuman dietary sialic acid.[PNAS]. *Proceedings of the National Acadamy of Sciences of the United States*, *100*(21), 12045–12050.

Tantamango-Bartley, Y., Jaceldo-Siegl, K., Fan, J., & Fraser, G. (2013). Vegetarian diets and the incidence of cancer in a low-risk population. *Cancer Epidemiology, Biomarkers & Prevention*, *22*(2), 286–294.

Toledo, E., Hu, F. B., Estruch, R., Buil-Cosiales, P., Corella, D., Salas-Salvado, J., & Martinez-Gonzalez, M. A. et al. (2013). Effect of the Mediterranean diet on blood pressure in the PREDIMED trial: Results from a randomized controlled trial. *BioMed Central Medicine*, *11*, 207.

Tonstad, S., Butler, T., Yan, R., & Fraser, G. E. (2009). Type of vegetarian diet, body weight, and prevalence of type 2 diabetes. *Diabetes Care*, *32*(5), 791–796.

Tonstad, S., Stewart, K., Oda, K., Batech, M., Herring, R. P., & Fraser, G. E. (2013). Vegetarian diets and incidence of diabetes in the Adventist Health Study-2. *Nutrition, Metabolism, and Cardiovascular Diseases*, *23*(4), 292–299.

Trichopoulou, A., & Lagiou, P. (1997). Healthy traditional Mediterranean diet: An expression of culture, history, and lifestyle. *Nutrition Reviews*, *55*(11), 383–389.

Tsunehara, C. H., Leonetti, D. L., & Fujimoto, W. Y. (1990). Diet of second-generation Japanese-American men with and without non-insulin-dependent diabetes. *The American Journal of Clinical Nutrition*, *52*(4), 731–738.

Tuomilehto, J., Lindstrom, J., Eriksson, J. G., Valle, T. T., Hamalainen, H., Ilanne-Parikka, P., & Uusitupa, M. et al. (2001). Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *The New England Journal of Medicine*, *344*(18), 1343–1350.

Turner-McGrievy, G. M., Barnard, N. D., Cohen, J., Jenkins, D. J., Gloede, L., & Green, A. A. (2008). Changes in nutrient intake and dietary quality among participants with type 2 diabetes following a low-fat vegan diet or a conventional diabetes diet for 22 weeks. *Journal of the American Dietetic Association*, *108*(10), 1636–1645.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Turner-McGrievy, G. M., Davidson, C. R., Wingard, E. E., & Billings, D. L. (2014). Low glycemic index vegan or low-calorie weight loss diets for women with polycystic ovary syndrome: A randomized controlled feasibility study. *Nutrition Research (New York, N.Y.)*, *34*(6), 552–558.

Turner-McGrievy, G. M., Davidson, C. R., Wingard, E. E., Wilcox, S., & Frongillo, E. A. (2015). Comparative effectiveness of plant-based diets for weight loss: A randomized controlled trial of five different diets. *Nutrition (Burbank, Los Angeles County, Calif.)*, *31*(2), 350–358.

United States Department of Agriculture (USDA) and United States Department of Health and Human Services. (USDHHS). (2010). Dietary Guidelines for Americans, 2010. Washington DC: USDA and USDHHS. Retrieved September 9, 2015, from http://health.gov/dietaryguidelines/dga2010/dietaryguidelines2010.pdf

United States Department of Agriculture (USDA) and United States Department of Health and Human Services. (USDHHS). (2015). Scientific Report of the 2015 Dietary Guidelines Advisory Committee. United States Department of Health & Human Services. Retrieved September 19, 2015, from http://health.gov/dietaryguidelines/2015-scientific-report/pdfs/scientific-report-of-the-2015-dietary-guidelines-advisory-committee.pdf

Uribarri, J., Cai, W., Peppa, M., Goodman, S., Ferrucci, L., Striker, G., & Vlassara, H. (2007). Circulating glycotoxins and dietary advanced glycation endproducts: Two links to inflammatory response, oxidative stress, and aging. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, *62*(4), 427–433.

Uribarri, J., Woodruff, S., Goodman, S., Cai, W., Chen, X., Pyzik, R., & Vlassara, H. et al. (2010). Advanced glycation end products in foods and a practical guide to their reduction in the diet. *Journal of the American Dietetic Association*, *110*(6), 911–916.

van Puyvelde, K., Mets, T., Njemini, R., Beyer, I., & Bautmans, I. (2014). Effect of advanced glycation end product intake on inflammation and aging: A systematic review. *Nutrition Reviews*, *72*(10), 638–650.

Vang, A., Singh, P. N., Lee, J. W., Haddad, E. H., & Brinegar, C. H. (2008). Meats, processed meats, obesity, weight gain and occurrence of diabetes among adults: Findings from Adventist Health Studies. *Annals of Nutrition & Metabolism*, *52*(2), 96–104.

Vergnaud, A. C., Norat, T., Romaguera, D., Mouw, T., May, A. M., Travier, N., & Peeters, P. H. et al. (2010). Meat consumption and prospective weight change in participants of the EPIC-PANACEA study. *The American Journal of Clinical Nutrition*, *92*(2), 398–407.

Vessby, B. (2000). Dietary fat and insulin action in humans. *The British Journal of Nutrition*, *83*(1), S91–S96.

Wang, Y., & Beydoun, M. A. (2009). Meat consumption is associated with obesity and central obesity among US adults. *International Journal of Obesity*, *33*(6), 621–628.

Wang, Z., Klipfell, E., Bennett, B. J., Koeth, R., Levison, B. S., Dugar, B., & Hazen, S. L. et al. (2011). Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature*, *472*(7341), 57–63.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Weggemans, R. M., Zock, P. L., & Katan, M. B. (2001). Dietary cholesterol from eggs increases the ratio of total cholesterol to high-density lipoprotein cholesterol in humans: A meta-analysis. *The American Journal of Clinical Nutrition*, *73*(5), 885–891.

West, R. O., & Hayes, O. B. (1968). Diet and serum cholersterol levels. A comparision between vegetarians and nonvegetarians in a Seventh-day Adventist group. *The American Journal of Clinical Nutrition*, *21*(8), 853–862.

White, D. L., & Collinson, A. (2013). Red meat, dietary heme iron, and risk of type 2 diabetes: The involvement of advanced lipoxidation endproducts. *Advances in Nutrition*, *4*(4), 403–411.

Willett, W. C. (2012). Dietary fats and coronary heart disease. *Journal of Internal Medicine*, *272*(1), 13–24.

Willett, W. C., & Stampfer, M. J. (2013). Current evidence on healthy eating. *Annual Review of Public Health*, *34*, 77–95.

Willett, W. C., Stampfer, M. J., Manson, J. E., Colditz, G. A., Speizer, F. E., Rosner, B. A., & Hennekens, C. H. et al. (1993). Intake of trans fatty acids and risk of coronary heart disease among women. *Lancet*, *341*(8845), 581–585.

Wong, J. M. W. (2014). Gut microbiota and cardiometabolic outcomes: Influence of dietary patterns and their associated components. *The American Journal of Clinical Nutrition*, *100*(1), 369S–377S.

World Cancer Research Fund (WCRF). (2007). Our Cancer Recommendations - Animal Foods. Retrieved February, 2015, from http://www.wcrf.org/int/research-we-fund/cancer-prevention-recommendations/animal-foods

World Cancer Research Fund. (2014). Our Cancer Prevention Recommendations. Retrieved November 17, 2014, from http://www.wcrf.org/int/research-we-fund/our-cancer-prevention-recommendations

World Health Organization (WHO). (2004). Global Strategy of Diet, Physical Activity and Health. Geneva, Switzerland: WHO. Retrieved September 10, 2015, from http://www.who.int/dietphysicalactivity/strategy/eb11344/strategy_english_web.pdf

World Health Organization (WHO). (2011). Global status report on noncommunicable diseases 2010. Geneva, Switzerland: WHO. Retrieved Septemver 10, 2015, from http://www.who.int/nmh/publications/ncd_report2010/en/

World Health Organization (WHO). (2014a). Cancer (Fact Sheet). Retrieved November 17, 2014, from http://www.who.int/mediacentre/factsheets/fs297/en/

World Health Organization (WHO). (2014b). Cardiovascular diseases (Fact sheet). Retrieved November 17, 2014, from http://www.who.int/mediacentre/factsheets/fs317/en

World Health Organization (WHO). (2014c). Diabetes (Fact sheet). Retrieved November 17, 2014, from http://www.who.int/mediacentre/factsheets/fs312/en/

Xu, J., Yang, X. X., Wu, Y. G., Li, X. Y., & Bai, B. (2014). Meat consumption and risk of oral cavity and oropharynx cancer: A meta-analysis of observational studies. *Public Library of Science (PLoS). One*, *9*(4), 1–9.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Xu, X., Yu, E., Gao, X., Song, N., Liu, L., Wei, X., & Fu, C. et al. (2013). Red and processed meat intake and risk of colorectal adenomas: A meta-analysis of observational studies. *International Journal of Cancer*, *132*(2), 437–448.

Xue, X. J., Gao, Q., Qiao, J. H., Zhang, J., Xu, C. P., & Liu, J. (2014). Red and processed meat consumption and the risk of lung cancer: A dose-response meta-analysis of 33 published studies. *International Journal of Clinical and Experimental Medicine*, *7*(6), 1542–1553.

Yang, W., Li, B., Dong, X., Zhang, X. Q., Zeng, Y., Zhou, J. L., & Xu, J. J. et al. (2014). Is heme iron intake associated with risk of coronary heart disease? A meta-analysis of prospective studies. *European Journal of Nutrition*, *53*(2), 395–400.

Yang, W. S., Wong, M. Y., Vogtmann, E., Tang, R. Q., Xie, L., Yang, Y. S., & Xiang, Y. B. et al. (2012). Meat consumption and risk of lung cancer: Evidence from observational studies. *Annals of Oncology*, *23*(12), 3163–3170.

Yokoyama, Y., Nishimura, K., Barnard, N. D., Takegami, M., Watanabe, M., Sekikawa, A., & Miyamoto, Y. et al. (2014). Vegetarian diets and blood pressure: A meta-analysis. *Journal of the American Medical Association Internal Medicine*, *174*(4), 577–587.

Zeuschner, C. L., Hokin, B. D., Marsh, K. A., Saunders, A. V., Reid, M. A., & Ramsay, M. R. (2013). Vitamin B12 and vegetarian diets. *The Medical Journal of Australia*, *199*(4), S27–S32.

Zhang, C., Schulze, M. B., Solomon, C. G., & Hu, F. B. (2006). A prospective study of dietary patterns, meat intake and the risk of gestational diabetes mellitus. *Diabetologia*, *49*(11), 2604–2613.

Zhao, Z., Li, S., Liu, G., Yan, F., Ma, X., Huang, Z., & Tian, H. (2012). Body iron stores and heme-iron intake in relation to risk of type 2 diabetes: A systematic review and meta-analysis. *Public Library of Science (PLoS). One*, *7*(7), 1–17.

Zheng, W., & Lee, S. A. (2009). Well-done meat intake, heterocyclic amine exposure, and cancer risk. *Nutrition and Cancer*, *61*(4), 437–446.

Zhu, H., Yang, X., Zhang, C., Zhu, C., Tao, G., Zhao, L., & Sun, X. et al. (2013). Red and processed meat intake is associated with higher gastric cancer risk: A meta-analysis of epidemiological observational studies. *Public Library of Science (PLoS). One*, *8*(8), 1–10.

Zhu, H. C., Yang, X., Xu, L. P., Zhao, L. J., Tao, G. Z., Zhang, C., & Sun, X. C. et al. (2014). Meat consumption is associated with esophageal cancer risk in a meat- and cancer-histological-type dependent manner. *Digestive Diseases and Sciences*, *59*(3), 664–673.

KEY TERMS AND DEFINITIONS

Advanced Glycation End-Products (AGEs): AGEs are formed when foods high in fat and protein (particularly meat) are cooked at high temperatures resulting in the browning or charring of food.

Haem Iron: The form of iron found in animal products including red and processed meats, which is readily absorbed by the body; a catalyst or pro-oxidant in many cellular reactions, including the endogenous formation of n-nitroso compounds.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Heterocyclic Amines (HCA): Chemicals formed in meat as a result of cooking meat at high temperatures. HCAs are known carcinogens.

N-Glycolylneuraminic Acid (Neu5Gc): A sialic acid (mainly in red meat) that promotes chronic inflammation and development of chronic diseases.

N-Nitroso Compounds: Carcinogenic substances formed in the gut following ingestion of haem iron (in red meat) and preservatives nitrites and nitrates (added to processed meats).

Polycyclic Aromatic Hydrocarbons (PAH): Chemicals formed when meat is charred during the cooking process. PAHs are also found in exhaust fumes and cigarette smoke and are known carcinogens.

Saturated Fat: A non-essential fatty acid found mostly in animal products, including red and processed meats and dairy, as well as coconut oil/milk, palm kernel oil and palm oil.

Trans-Fats: A non-essential fatty acid formed during the partial hydrogenation of vegetable oils to improve stability and shelf life; small amounts are naturally occurring in ruminant animals.

Trimethylamine-N-Oxide (TMAO): A toxic compound associated with increased cardiovascular disease risk, produced by the action of gut bacteria on L-carnitine and choline.

This work was previously published in Impact of Meat Consumption on Health and Environmental Sustainability edited by Talia Raphaely and Dora Marinova, pages 131-177, copyright year 2016 by Information Science Reference (an imprint of IGI Global).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 8
HPV Detection Methods:
Towards Personalized Prevention

Aris Spathis
University of Athens, Greece

Abraham Pouliakis
University of Athens, Greece

Christine Kottaridi
University of Athens, Greece

Stavros Archondakis
401 Army General Hospital, Greece

Petros Karakitsos
University of Athens, Greece

ABSTRACT

Human papilloma viruses (HPVs) have been acknowledged to be the leading risk factor of cervical intra-epithelial lesion creation (CIN) and cervical cancer development (CxCa). Many different techniques have been created and utilized in HPV detection and monitoring with a vast amount of them being commercialized and few of them integrated in official screening strategies. A growing trend for DNA typing of the 14 most commonly accepted high risk HPV types has been introduced, supporting that in many cases molecular testing could replace classic morphologic diagnostic routines, even though DNA detection has lower specificity than other molecular and morphology tests. However, there have been limited attempts in combining data from all different techniques to provide efficient patient triaging schemes, since, apart from the obvious increase of patient cost, the amount of data and its interpretation in patient management has been impossible. Complex computer based clinical support decision systems, many of which are based on artificial intelligence may abolish these limitations.

INTRODUCTION

Human papilloma viruses (HPV) are members of the papillomaviridae family. HPVs are small (7,000-8,500bp) double stranded DNA viruses with no RNA phase in their life cycle that can only be reproduced in keratinocytes. Over 100 different types have been identified and isolated, both in humans and animals (de Villiers, Fauquet, Broker et al, 2004). Around 40 types have been isolated from human cervical specimens and HPVs have been classified, according to their epidemiology in cervical lesions, as low

DOI: 10.4018/978-1-5225-0549-5.ch008

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

risk and high risk (Dunne, Unger, Sternberg et al, 2007; Munoz, Bosch, Sanjose et al, 2003; Trottier & Franco, 2006). Although HPVs have been mostly isolated from cervical samples they have also been detected in samples from the anus, vulva, vagina, skin and even head and neck cancers.

HPVs are small double-stranded DNA viruses encoding ten proteins in total, out of which two (E6 and E7) are considered oncoproteins (Münger, Basile, Duensing et al, 2001; Mantovani & Banks, 2001) and two (L1 and L2) are the structural proteins of the viral capsid. The rest proteins include one helicase (E1), one transcription factor (E2), two proteins that induce host proliferation (E5) and viral replication (E4) and a short protein that is expressed only in few papillomavirus types that may substitute for the E6 open reading frame. E6 has been shown to interact with the p53 tumor suppressor protein leading to its degradation, while E7 inactivates members of the pRB family of tumor suppressor proteins and activates telomerases. Both proteins enhance cell proliferation and down regulate apoptosis.

HPVs proteins are named after the stage of the viral infection cycle; E proteins are early expressed proteins and are responsible for the replication of the virus, while L proteins are late expressed proteins after the virus has replicated. HPVs infect cells of the basal epithelium of the cervix via minor abrasions of the squamous epithelium. At first the virus reproduces as an extra-chromosomal plasmid during cell division and these infections are sub-clinical. Later on and after the expression of viral oncogenes, the virus reproduces by thousands and infection spreads in more cells (activated infection). After a long lasting infection, the virus is proposed to be integrated in the host genome leading to deregulation of gene expression, disruption of tumor suppressor genes and the accumulation of genetic changes that ultimately result in malignant transformation of epithelial cells (Andersson, Safari, Mints et al, 2005). A persistent infection by HPV aided by other parameters results in HPV integration in host DNA leading to progression to high grade lesions (Wallin, Wiklund, Angström et al, 1999). Most HPV infections are cleared by the host's immune system before they progress to the creation of intra-epithelial lesions. Only persistent infection with high risk HPV types has been acknowledged as the most important risk factor for cervical cancer development (Walboomers, Jacobs, Manos et al, 1999).

First connection of HPVs with cancer was proposed by Stefania Jablonska in Poland in 1972 in epidermodysplasia verruciformis (Jabłońska, Orth, Jarzabek-Chorzelska et al, 1978). Initial evidence supporting that theory emerged in 1978, when Jablonska and Gerard Orth discovered HPV-5 in skin cancer. In 1976 Harald zur Hausen published the hypothesis that human papillomavirus plays an important role in the cause of cervical cancer, and finally in 1983 and 1984 zur Hausen and his collaborators identified HPV16 and HPV18 in cervical cancer (zur Hausen, 1976). HPVs' linking to cervical carcinogenesis has led to the award of the 2008 Nobel Prize of medicine to Dr. Herald zur Hausen for his pioneering work (zur Hausen, 1976, 1991).

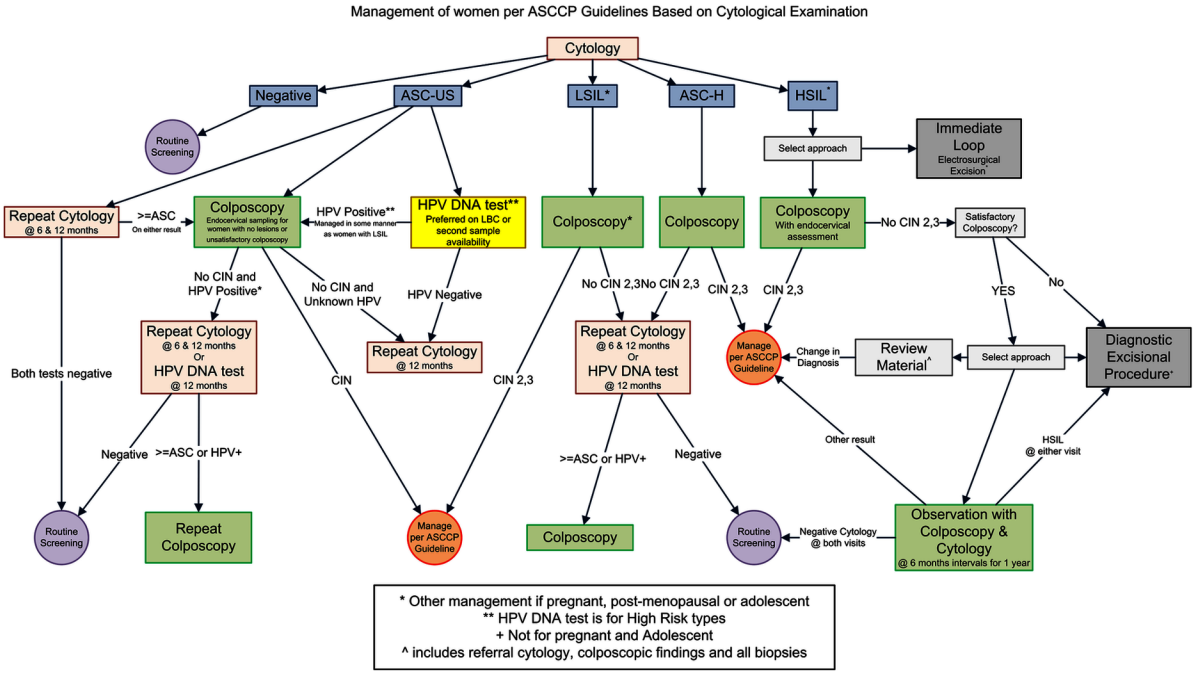
HPV infection prevalence can vary substantially across different age groups, study populations and countries (Revzina & Diclemente, 2005). Although the prevalence of HPV is about the same worldwide and Cervical Intraepithelial Neoplasias (CIN) develop gradually resulting in progression to CxCa within a time frame between 10-20 years, the worldwide incidence of CxCa differs. This reflects different strategies countries have developed, mainly due to the adaptation of organized screening programs based on periodical cytological examination of cervical smears. This examination was invented by a Greek medical doctor, George Papanicolaou and is performed via microscopic morphological evaluation of cervical cell alterations according to a grading system. The most commonly used system is the Bethesda 2001 (Solomon, Davey, Kurman et al, 2002). Abnormal findings during the microscopic inspection of cells results, depending of the severity of the findings, in either repeated testing after 6-12 months or immediate referral to colposcopy. Should a visible lesion be identified a biopsy is taken followed by a

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

therapeutic surgical intervention after histological verification (Wright, Massad, Dunton et al, 2007). The summarized triaging options guidelines from the American Society for Colposcopy and Cervical Pathology (ASCCP) are presented in Figure 1. Organized screening programs have reduced substantially CxCa incidence and fatality rates (Peto, Gilham, Fletcher et al, 2004). However, CxCa remains a serious problem of public health worldwide, ranked as the third most common cancer worldwide among women, and the third leading cancer related cause of death with over 500,000 and 250,000 estimated new cases and deaths, respectively (Jemal, Bray, Center et al, 2011). Even though HPV vaccination, introduced in 2006 in two commercial forms (Gardasil, marketed by Merck and Cervarix, marketed by GlaxoSmithKline), is thought to further reduce HPV related CxCa development (Kahn, 2009), screening and treatment guidelines have not changed for women that have been vaccinated (Harper, 2009).

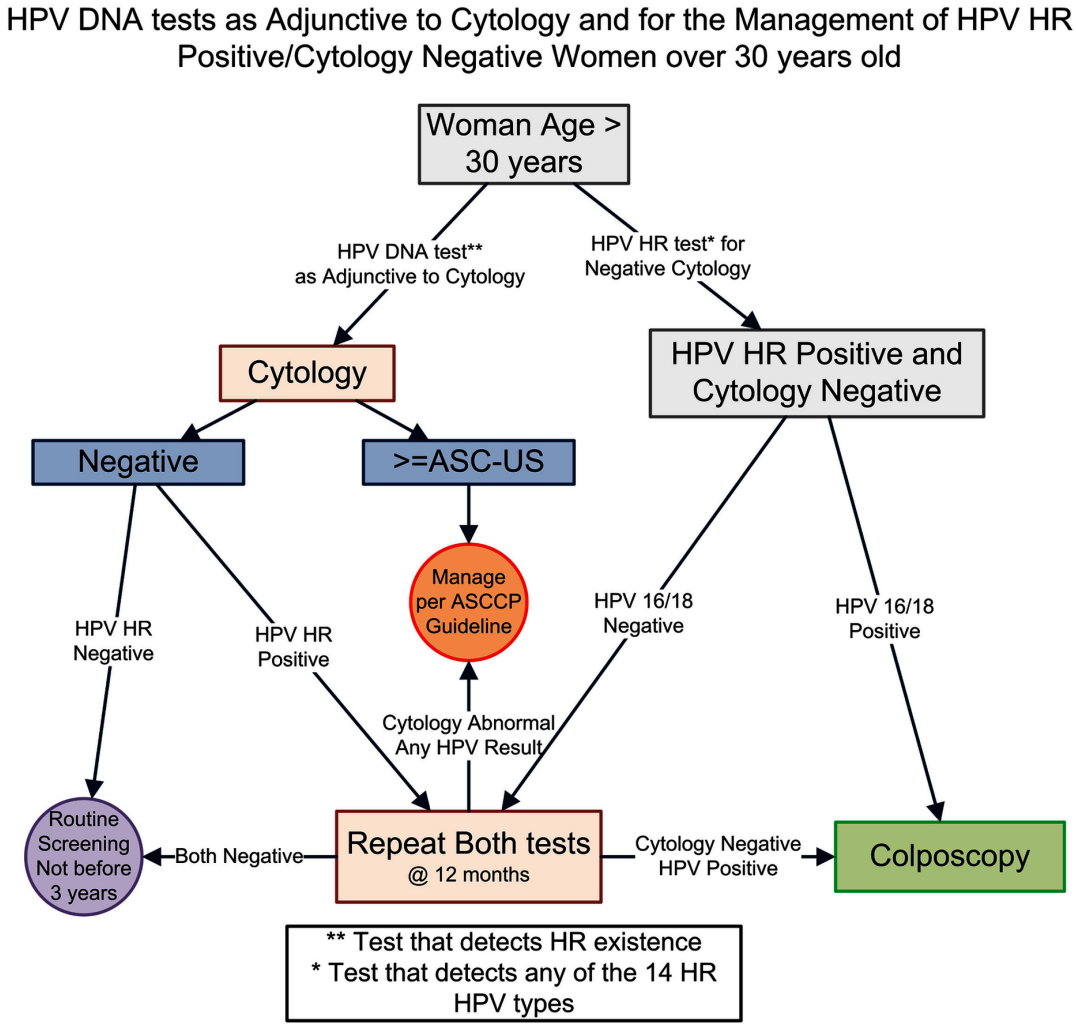
More recently molecular biology tests have been added as ancillary techniques in order to identify HPV infections, aid in the screening and triaging options available for clinicians, and to identify cases where morphology was insufficient (Figure 2). A milestone in that direction was the introduction of Liquid Based Cytology (LBC) that allows cervical samples to be stored in a medium which preserves cell morphology and biological macromolecules, and also allows ancillary testing to be executed from the residual cell suspension, eliminating the need for a recall visit. Two major commercial LBC mediums are available, ThinPrep® by Hologic® and Surepath™ by Becton Dickinson. Both types offer systems that perform monolayer slide preparations and both are alcohol based; the first methanol and the latter ethanol, both of which have been shown to be excellent preservatives of nucleic acids. LBC has limited the number of insufficient samples, but there is still controversy as to the increase of diagnostic accu-

Figure 1. Monitoring and triaging options based on cytology findings, staged according to the Bethesda system. Notice that HPV DNA testing has been introduced: a) in ASC-US category, that is the cytologic result with the highest degree of uncertainty and b) as a reflex test in retesting LSIL and ASC-H.



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 2. Monitoring and triaging options of women >30 with HPV testing. A high risk HPV positive result with HPV16/18 will be sent for Colposcopy, while any other high risk positive result is treated as an AS-US. Combination of negative cytology and HPV allows the increase of the screening interval to three years.



racy compared to conventional cytology (Siebers, Klinkhamer, Grefte et al, 2009). As a result of LBC introduction, many companies around the world have developed commercially available kits that aim in HPV detection utilizing different strategies.

The aim of this chapter is the presentation of the techniques that are available in HPV detection, the technology they utilize, their drawbacks and their advantages over other ones. Finally, we would like to point out the potential problems that arise by having a vast arson of molecular data that at present cannot be evaluated by clinical physicians. Furthermore, we present three modern methods based on artificial intelligence or statistical techniques, these methods provide encouraging results for the exploitation of the multitude of available parameters for the triage of women and the prognosis of CIN.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

HPV DETECTION TECHNIQUES

HPV detection techniques vary both in the molecule they identify, their accuracy and in their workflow implementation efficiency. Since microscopic examination of cervical cells can only identify the morphological changes that are acquired, one can only assume the presence of the virus. Even though initial evidence of HPVs' presence were acquired with electron microscopy, via identification of the free viral particles, implementation of such a method in large scale is uneconomical and workflow inefficient. Since HPV is a double stranded DNA virus, most early and current techniques have targeted viral DNA identification with an amplification step included, since viral load if often low. That step can be either a DNA amplification step or a signal amplification step. After first evidence on the viral cycle of HPVs arose and the E6 and E7 genes were characterized as viral oncogenes, mRNA detection became a new trend in HPV detection. Finally, the virus has a protein outer capsid that surrounds it during the release from an infected cell and some efforts have been made in that protein identification. Due to numerous reasons analyzed later on, serological identification of HPV has been scarcely used.

HPV DNA DETECTION

Detection of HPV DNA is performed either in cervical cytology swabs or biopsies. Liquid based cytology is the most widely used starting material, since nucleic acid preservation is far superior to conventional cytology samples. Most techniques are based on the detection after some type of amplification of the L1 gene of HPV, while those that offer partial or full typing are based on hybridization of viral DNA to specific probes.

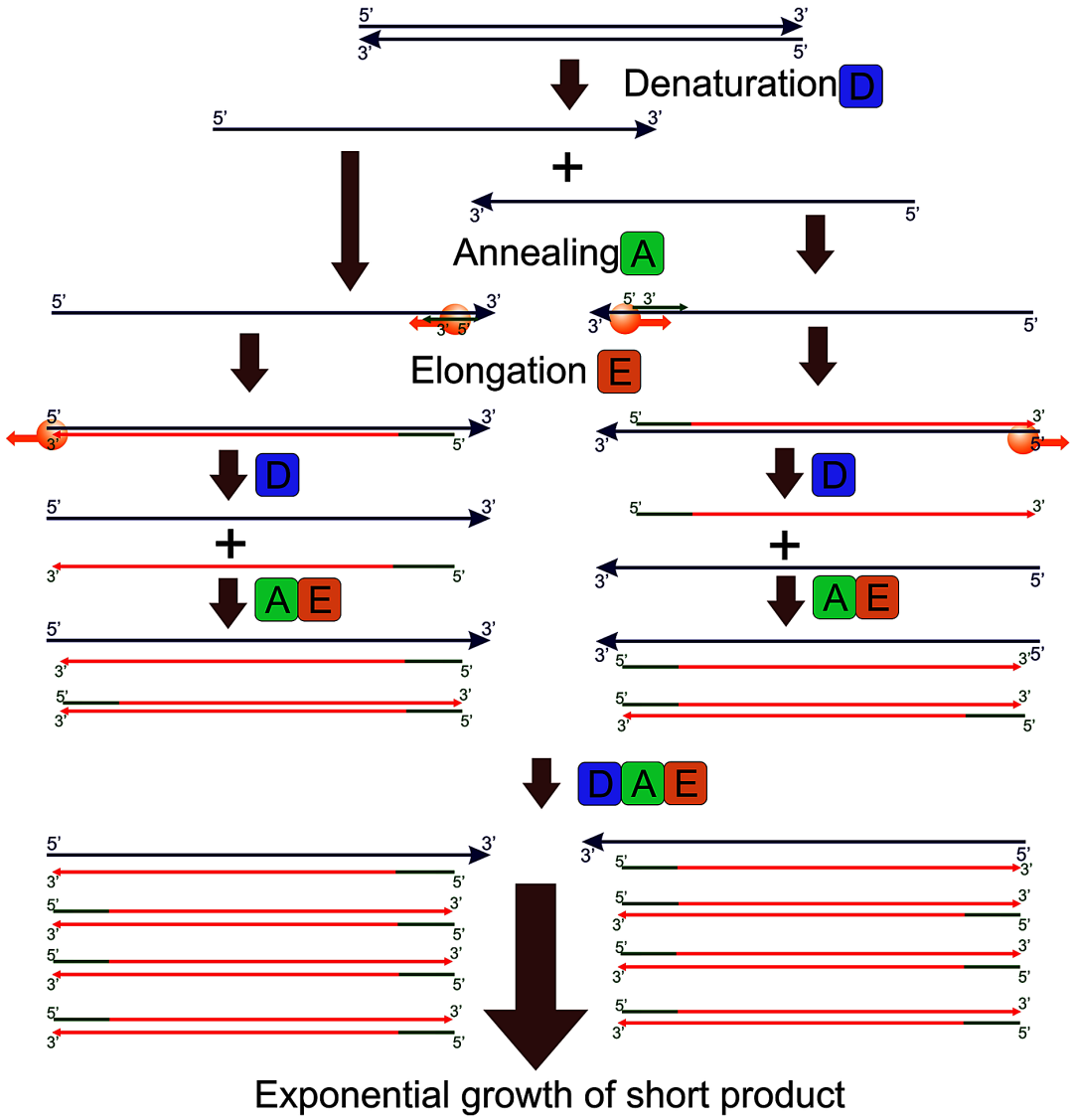
PCR Based Techniques

Polymerase Chain Reaction (PCR) is a molecular biology technique inspired by Kary Mullis (Bartlett & Stirling, 2003) that has led to an exponential growth of molecular biology techniques for viral identification and characterization. This technique utilizes a thermo stable DNA polymerase that, in the presence of a pair of 15-25 base long single strand DNA nucleotides (primers) and a starting template at which they anneal to, is able to amplify a selected DNA sequence. One of the two primers is complementary to the start of the amplified sequence and the other is complementary to the reversed end of the end of the sequence (Figure 3). Amplification occurs in a PCR machine also known as thermo cycler, since it is a machine that performed repeated cycles of incubation at specific temperatures. A common PCR program includes: a) an initial template denaturation step, where the DNA template becomes single stranded, b) 30-40 cycles of bi) denaturation of template DNA and PCR products, bii) annealing of primers to the DNA template and PCR products, biii) elongation of the annealed DNA sequences and finally c) a final elongation step to fully elongate all sequences. A PCR program of 35 cycles results, from a single starting template copy and with a theoretical amplification efficiency of 100%, in the creation of 2^{35} copies that equals to 34.3 billion copies.

PCR based techniques have high sensitivity, but usually suffer from false positives due to cross-contaminations and miss-priming. Furthermore, overall accuracy for HPV detection varies greatly depending on the primer set, the reaction conditions and the enzyme used. Specifically for HPVs there were two strategies used; one that utilizes primers that are type specific resulting in increased specificity and one that uses primers that are designed for well conserved sequences of a target gene, usually L1,

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 3. The Polymerase Cycle Reaction. The denaturation step (94-95°C) is used to separate the two DNA strands (blue), the annealing (45-60°C) to allow the hybridization of the primers (green), and the extension step (60-72°C) to allow the polymerase (orange sphere) to extend the two strands. In each cycle, if an ideal amplification efficiency of 100% is achieved, the original sequence is amplified by 2^n, where n is the number of cycles.



resulting in amplification of various HPV types with a single primer set. The latter has been widely used in detecting HPVs in cervical samples and has provided clinical evidence of the connection of HPVs with cervical cancer. The three primer sets that are widely used and amplify part of the L1 viral gene are: the GP5/6 and their elongated version GP5+/6+ primer sets (Gravitt, Peyton, Alessi et al, 2000), the degenerate MY09/11 and their un-degenerate counterpart PGMY (Manos, Kinney, Hurley et al, 1999) and the SPF10 primers set (Kleter, van Doorn, Schrauwen et al, 1999). These primer sets have a final PCR product ranging from 64bp to 450bp that is visualized after electrophoresis on an agarose gel.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

While most primary studies in HPV detection used PCR in order to identify viral DNA presence, current commercially available kits utilize a PCR step prior to detection of HPV types with immobilized probes. Their primary advantage is enhanced sensitivity due to the amplification step and the ability to type the HPV present, contributing significantly in epidemiologic data collection of HPV infections.

AMPLICOR® - LINEAR ARRAY HPV GENOTYPING Tests

These are basically two tests, the first for HPV detection and the latter for HPV typing created and marketed by Roche Molecular Diagnostics (Roche, Switzerland) and both under FDA review for approval. The first step is common in both tests allowing DNA extraction, validated from samples in ThinPrep® or Surepath®, to be performed once for both tests. The next step is a PCR using primers that amplify a 165bp sequence of L1 gene and a 268bp sequence of human beta-globin, used as internal control for the AMPLICOR® HPV test. PGMY09/11 are used for HPV detection in the LINEAR ARRAY. The final step is hybridization with immobilized probes in either an ELISA plate, for the AMPLICOR® HPV, or on strip LINEAR ARRAY, followed by the colorimetric detection using a biotin-streptavidine-peroxidase sandwich technique. These tests allow the detection of 13 high risk HPV types (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59 and 68) and the detection of infections and co-infections of up to 35 HPV genotypes both high risk (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68, 73 [MM9], 82[MM4]) and low risk (6, 11, 26, 40, 42, 53, 54, 55, 61, 62, 64, 67, 69, 70, 71, 72, 81, 83 [MM7], 84 [MM8], IS39, CP6108). On a note, LINEAR ARRAY cannot type co-infection of HPV52 with HPV58, while analytical limits tested with plasmids revealed a limit of detection (LOD) that varied from 76-20,000 copies/ml, depending on the type tested. On the other hand LOD for AMPLICOR® HPV were significantly lower ranging from 100-240 copies, depending on the type tested (Table 1).

Clinical utility of both tests and their newer variation that utilizes the COBAS platform were the target of a large clinical study (ATHENA) and are summarized in two major publications (Castle, Stoler, Wright et al, 2011; Stoler, Wright, Sharma et al, 2011). The first study compares the tests with LBC, while the latter with Hybrid Capture 2 for samples with cytology result Atypical Squamous Cells of Unknown Significance (ASC-US). Their results support that DNA testing, especially the COBAS that detects high risk HPV types with individual HPV16 and HPV18 identification, can detect CIN3+ cases with increased sensitivity compared to cytology. Both propose that the COBAS HPV test is a valid method for ASC-US triaging.

CLART® HPV2

The Clinical Arrays HPV2 kit (Genomica, Spain) uses the biotinilated PGMY version of the general primers to amplify a conserved region of L1 gene and typing of 35 distinct HPV types is performed with hybridization on an array tube,that is basically a low density micro-array printed on the bottom of a tube. There are two internal controls; beta globin for efficiency of DNA extraction and an internal control for amplification efficiency. The biotinilated PCR products are hybridized with the immobilized probes and a peroxidase – streptavidine complex is added that converts the g-dianisidine to an insoluble product. Detection of the precipitate is observed on sites that hybridization has occurred. This technique allows detection of infections and co-infections of up to 35 HPV genotypes that have been classified as high-risk (16, 18, 26, 31, 33, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 68, 70, 73, 82 and 85) and low risk (6, 11, 40, 42, 43, 44, 54, 61, 62, 71, 72, 81, 83, 84 and 89), according to Dune et al. (Dunne et al., 2007). The test has been validated from various starting materials including Liquid Based Cytology, swabs and

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 1. Technology, limit of detection and clinical efficiency of most widely used commercial HPV detection kits

Test	Hybrid Capture 2	Cervista®	cobas® 4800 HPV	Linear Arrays HPV	CLART2 HPV	NucliSens EasyQ HPV	Oncotect™ E6, E7 mRNA	Aptima HPV Assay
Company	Qiagen	Hologic	Roche	Roche	Genomica	Biomerieux	InCellDx	Gen-Probe
Approvals	FDA, CE-IVD	FDA, CE-IVD	FDA, CE-IVD	CE-IVD	CE-IVD	CE-IVD	CE-IVD	FDA, CE-IVD
HPV Detection	DNA	DNA	DNA	DNA	DNA	mRNA	Intracellular mRNA	mRNA
Detection Technology	Chemiluminescence (RNA:DNA)	Invader® (Fluorescent probes)	Fluorescent probes	Hybridization in Linear strips	Hybridization in Array tubes	Fluorescent beacons	Fluorescent probes	HPA
Detection Organ	Luminometer	Fluorometer	Real-time PCR	Colorimetric	Colorimetric	Real-time PCR	Flow Cytometry	Luminometer
HPVs Detected	13HR + 5 LR	14 HR	14 HR	16 HR + 21 LR	20 HR+ 15 LR	5 HR	HR	14 HR
Typing	No	HPV16, 18	HPV16, 18	Yes*	Yes	HPV16,18,31,33,45	No	No
Extraction	DNA/RNA Manual	DNA Manual or automated	DNA Automated	DNA Manual or Automated	DNA	DNA/RNA Semi- or Fully Automated	None	mRNA Fully Automated
Amplification	None	Invader®	PCR	PCR	PCR	NASBA	None	TMA
Result Interpretation	Automated	Automated	Automated	Manual	Automated	Automated	User cut-off	Automated
Internal controls for	Human DNA	Human DNA	Human DNA	Human DNA	Amplification+DNA	Human mRNA	Fixed cell lines	Yes
Sensitivity (CIN2+)	87.2-96.3	95.5	90-95.2	91-93.3 [5,65]	94-96.9	74.1-85.7	81.4-84	97.7
Specificity (CIN2+)	19.5-71.1	63.2	24.0-70.5	48.1-50.8	33.9-71.9[5]	70.8-84.9 [1,7,8]	74-91	28.8-52.9
PPV (CIN2+)	13.7-37.4	10.1	14.0-37.6	14.6-72.9	28.7-87	55.4-64.4	48-78	32.1-39.3
NPV (CIN2+)	99.1	96.9	99.2	79.5-98.7	93.1-97.6 [55]	81.2-95.1	82.8-93.1	96.3
Limit of detection	50,000 copies/ml	1,250-7,500 copies/rxn	150-1,200 copies/ml	76-20,000 copies/ml	10-100 copies/rxn	500-5000 copies/rxn	1.5-5% of cells	300 copies/rxn
Time / run	6-7 h	6-7 h	8 h	8-10h	8-10h	8 h	3h	4.5-6 h
Samples / run	90	28	288	24	24	28	24	100-250
Cost / Sample	40€	60€	60€	90€	90€	70€	50€	40€
References	(Castle et al., 2011; Stoler et al., 2011; Szarewski et al., 2008)*	(Johnson et al., 2008)*	(Castle et al., 2011; Stoler et al., 2011; Szarewski et al., 2008)*	(Gravitt et al., 2008; Stevens, Garland, Rudland et al, 2007)	(Chranioti, Spathis, Aga et al, 2012; Pista,Verdasca & Oliveira, 2011; Rebolj, Preisler, Ejegod et al, 2014)	(Keegan, Mc Inerney, Pilkington et al, 2009; Molden, Kraus, Karlsen et al, 2006; Szarewski et al., 2008) *	(Coquillard,Palao & Patterson, 2011; Kottaridi, Georgoulakis, Kassanos et al, 2010; Spathis, Kottaridi, Chranioti et al, 2012)	(Castle et al., 2011; Szarewski et al., 2008)

HPA: Hybridization Protection Assay, NASBA: Nucleic Acid Based Amplification, TMA: Transcritpion Mediated Amplification,
Prices are estimates from market inquiries in Greece *Results are from studies in ASC-US samples

paraffin embedded tissues. A main difference this test has compared to similar ones is that all probes for HPVs and controls are printed in triplicate and analysis of the final image is performed in an automated manner using a proprietary image system. This fact eliminated subjectivity caused by optical estimation of colorimetric strip tests, while the replicates allow verification of the sediment creation. The analytical sensitivity for most HPV types is between 100 and 10 copies of viral genome equivalents (Table 1).

Due to the tests ability to type 35 HPV types it has been used in epidemiological studies to identify the prevalence of HPV infection in different populations (Pista, de Oliveira, Cunha et al, 2011; Tsiodras, Hatzakis, Spathis et al, 2011). When compared to other HPV detection techniques, CLART® HPV2 has shown variable results from very good (Pista, Oliveira, Verdasca et al, 2011) to mediocre (García-Sierra, Martró, Castellà et al, 2009; Szarewski, Ambroisine, Cadman et al, 2008). Disagreement with HC2 could be attributed to the false cross-reaction that has been shown in HC2 (Castle, Solomon, Wheeler et al, 2008; Terry, Ho, Londesborough et al, 2001), while in the study by Szarewski the result could be biased since CLART results were only available for 737 out of 953 samples.

INNO-LiPa HPV

The INNO-Lipa HPV test (Innogenetics, Belgium) is a reverse dot-blot hybridization based on colorimetric identification of the hybridized PCR product of the SPF10 primers (Kleter et al., 1999). The test is able to identify 28 HPV types categorized as known and probable high-risk HPV genotypes (16, 18, 26, 31, 33, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 68, 73, 82) as well as a number of low-risk HPV genotypes (6, 11, 40, 43, 44, 54, 70) and some additional types (69, 71, 74), but cannot allow discrimination between HPV69 and HPV71. HPV typing is performed on a strip where immobilized probes are present for all types. The test includes a human gene (HLA-DPB1) as an internal control of efficient DNA isolation and lack of PCR inhibition. Furthermore instead of adenine, uracil is used for the PCR and an enzyme (Uracil-N-glycosylase) is included in order to avoid cross-contamination from a previously amplified sample. Its' use has been validated against both Hybrid capture 2 and the Linear array HPV genotyping test on starting material stored in Surepath™ (Fontaine, Mascaux, Weyn et al, 2007). Since the PCR product of SPF10 is only 60-65bp long their efficiency in heavily degraded DNA, as is paraffin extracted DNA, is proposed to be superior to other PCR based techniques. A study that compared matched formalin fixed paraffin embedded (FFPE) biopsies to cervical scrapings found that in overall HPV positivity the INNO-LiPa test had a 100% concordance, while some types were lost either in the scraping or the biopsy reflecting more the importance of the sampling area rather than the inability to amplify HPVs' DNA from FFPE (Dal Bello, Spinillo, Alberizzi et al, 2009).

HPV Direct Flow-chip

HPV Direct Flow CHIP (Master Diagnóstica, Granada, Spain) is a new commercially available product for sensitive HPV detection and genotyping, CE-IVD marked, in compliance with European Union diagnostic medical device manufacturing standards. The direct PCR system allows [GP5+/GP6+]-based multiplex amplification from crude-cell extracts and does not require previous DNA extraction. The assay detects 18 high-risk or putative high-risk genotypes (16, 18, 26, 31, 33, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 68, 73, and 82MM4) and 18 low-risk genotypes (6, 11, 40, 42, 43, 44, 54, 55, 61, 62, 67, 69, 70, 71, 72, 81, 84, and 89). Biotinylated amplicons are hybridized to HPV CHIP membranes containing immobilized probes for hybridization control (beta-globin), HPV-consensus sequences, and genotype-specific HPV detection. The hybridization is performed automatically in sets of 15 samples using the

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

e-BRID System™ (Master Diagnóstica, Granada, Spain), which allows the DNA target molecules to cross the membrane and bind to the complementary probes. Colorimetric detection is carried out by adding NBT-BCIP substrate that reacts with the alkaline phosphatase (AP) coupled to the sequences, creating insoluble purple precipitates. The entire process, from the start to the final read-out, is completed in less than three hours that is a significant enhancement compared to many other typing kits (Herraez-Hernandez, Alvarez-Perez, Navarro-Bustos et al, 2013; Herraez-Hernandez, Preda, Alonso et al, 2013).

The analytical sensitivity of the test was validated after successful participation in the 2011 WHO HPV LabNet Proficiency Study (Eklund, Forslund, Wallin et al, 2014), in which the kit demonstrated 100% agreement with reference values in 43 samples evaluated at different concentrations (from 5 to 500 GE) in the setting of single and multiple infections. Given that it offers direct PCR from clinical specimens without a DNA purification step, this novel test may be a valuable tool for automated, rapid, and sensitive HPV genotyping, especially in large-scale vaccine surveillance and epidemiology studies.

PapilloCheck® HPV-Screening

PapilloCheck® and PapilloCheck® high-risk (Greiner Bio-One GmbH, FrickenhausenGermany) are both microarray-based diagnostic tools for the simultaneous detection and identification of different panels of HPV types. ThePapilloCheck® high-risk test kit is designed to meet the demands of primary cervical cancer screening by targeting 14 of the most carcinogenic high HPV types (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68). PapilloCheck® is a test kit for the genotyping of 24 HPV types, including 18 high HPV (16, 18, 31, 33, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 68, 70, 73, 82) and 6 low (6, 11, 40, 42, 43, 44) HPV types.

After the extraction of viral and human genomic DNA from a cervical smear, a 350 bp fragment of the E1 gene is amplified in the presence of a set of HPV-specific primers by polymerase chain reaction (PCR), in presence of dUTP for carry-over contaminations. To avoid false negative results, a fragment of the human "house-keeping gene" ADAT1 (Adenosine deaminase, tRNA-specific1) is amplified in the same reaction to verify the presence of human DNA. Consequently, potential carry-over contaminations from previous PCR reactions can be eliminated via UNG treatment. The amplification products are then hybridized to the on-chip controls and HPV-specific DNA-probes attached to the DNA-chip surface. Every DNA-chip contains 12 DNA-microarrays allowing the simultaneous analysis of 12 cervical samples. During hybridization, the bound DNA is also fluorescently labelled and unbound DNA is removed in the subsequent washing steps. Finally, the PapilloCheck® DNA-chip is automatically scanned, analyzed and evaluated using the CheckScannerTM and CheckReportTMSoftware, respectively.

The PapilloCheck® DNA-chip incorporates several control systems which monitor all critical steps of the assay and chip processing, e.g. DNA extraction, PCR and hybridisation, as well as spot homogeneity and printing quality. Thus, false negative and false positive results are virtually excluded rendering the analysis highly reliable.

Hybribio 21 HPV GenoArray

The Hybribio 21 HPV GenoArray test (Hybribio, Wanchai, Hong Kong) is an L1 consensus primer-based PCR assay and is capable of amplifying 21 HPV genotypes, including 13 high risk types (types 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, and 68), 2 probable high risk types (types 53 and 66), and 6 low-risk (types 6, 11, 42, 43, 44) and unknown-risk types (CP8304 [HPV-81]). The amplicon is subsequently denatured and subjected to hybridization. The assay utilizes a flowthrough hybridization technique by

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

actively directing the targeting molecules toward the immobilized probes within the membrane fibers, with the complementary molecules being retained by the formation of duplexes. After a stringent wash, the hybrids are detected by the addition of the streptavidin-horseradish peroxidase conjugate, which binds to the biotinylated PCR products, and after addition of the substrate (nitroblue tetrazolium-5-bromo-4-chloro-3-indolylphosphate) a purple precipitate is produced at the probe dot. The results are interpreted by direct visualization. The positive control provided with the kit and the two negative controls (HPV-negative C33-A cells and PCR water) are included in each set of PCRs to assess the performance of the test. After hybridization, the presence of a positive result for both the "internal control" and the "biotin" dots within the membrane indicate that the isolated DNA is of good quality, the enzyme conjugate is valid, and the hybridization process is proper.

The performance characteristics of the GenoArray genotyping assay were compared with the commercial Roche Linear Array HPV genotyping assay (Liu, Leung, Chan et al, 2010). The absolute agreement between the two assays was 95.9% and the kappa value of 0.63 indicated good agreement.

GenoFlow HPV array (DiagCor)

The GenoFlow HPV Array test kit (GenoFlow, DiagCor Bioscience, Hong Kong) is a PCR-based assay that co-amplifies a human gene as an internal control to measure the integrity of the sample. This assay, by employing a membrane-based flow-through hybridization technology, identifies 33 high risk and low risk HPV viral genotypes (6, 11, 16, 18, 26, 31, 33, 35, 39, 40, 42, 43, 44, 45, 51, 52, 53, 54, 55, 56, 57, 58, 59, 61, 66, 68, 70, 71, 72, 73, 81, 82, 84). A novel universal probe is also designed to capture HPV genotypes that were not included in the test panel in order to expand the viral coverage. The amplicons are genotyped using a "flow-through" technology which involves hybridization, enzyme conjugation and colorimetric development, which can be finished in 35 min. A sample will be considered valid only if the hybridization control (for monitoring success of hybridization), amplification control (for access of sample integrity or successful PCR reaction), and/or specific HPV type on the membrane are obtained. Positivity is shown by colored dots on the membrane, which is recorded by scanning of the membrane by flat based scanner.

The GenoFlow Test was evaluated by comparing the genotyping results of 100 samples with Roche Linear Array HPV Genotyping Test (Wong,Ching & Chow, 2010). The tests showed good agreement in detection of HPV-positive samples (concordance rate=95%, Cohen's Kappa=0.896), with good agreement in detection of HR HPVs (Cohen's Kappa=0.876). The GenoFlow Test showed high sensitivity (95%), high specificity (95%), low false positive rate (3.33%) and low false negative rate (7.50%). The GenoFlow Test, offers advantages of reduction in turnaround time and manpower.

Real-time PCR Based Techniques

More recently developed tests for HPV detection have started using the Real-time PCR platforms. Compared to conventional PCR, Real-time PCR has many advantages such as the lower detection limit, due to increased sensitivity, and the ability to use several chemistries that allow superb specificity. Increased sensitivity is lies in PCR product identification via fluorescence production, while the addition of Taqman probes that can be only be cleaved by a completely complementary sequence, enhances specificity. Furthermore, Real-time PCR methods have lower hand-on time, since detection ends with the PCR reaction.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

cobas® HPV Test

The COBAS HPV test, mentioned earlier, is the only qPCR test currently approved since April of 2011 from the U.S. Food and Drug Administration (FDA). A primer set pool is used to amplify a sequence of approximately 200 nucleotides within the polymorphic L1 region of the HPV genome from the 14 high-risk types (HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and 68) that have been considered to be the most oncogenic types (Bouvard, Baan, Straif et al, 2009). Fluorescent oligonucleotide probes bind to polymorphic regions within the sequence defined by these primers. An additional primer pair and probe target the human beta globin gene amplifying a 330bp region to provide a process control. The test has been validated on ThinPrep aliquoted samples without a slide preparation prior to testing (Table 1).

Clinical evaluation of cobas was performed in a large clinical study (ATHENA), the results of which have been summarized in two major publications (Castle et al., 2011; Stoler et al., 2011). The first study compares the tests with LBC, while the latter with Hybrid Capture 2 for samples with cytology result Atypical Squamous Cells of Unknown Significance (ASC-US). Their results support that DNA testing, especially the COBAS that detects high risk HPV types with individual HPV16 and HPV18 identification, can detect CIN3+ cases with increased sensitivity compared to cytology. Both propose that the COBAS HPV test is a valid method for ASC-US triaging.

In the original FDA approval of the cobas the indented use was: “a) to screen patients 21 years and older with ASC-US cervical cytology test results to determine the need for referral to colposcopy, b) to be used in patients 21 years and older with ASC-US cervical cytology results, to assess the presence or absence of high-risk HPV genotypes 16 and 18. This information, together with the physician’s assessment of cytology history, other risk factors, and professional guidelines, may be used to guide patient management. The results of this test are not intended to prevent women from proceeding to colposcopy, c) in women 30 years and older, the cobas® HPV Test can be used with cervical cytology to adjunctively screen to assess the presence or absence of high risk HPV types. This information, together with the physician’s assessment of cytology history, other risk factors, and professional guidelines, may be used to guide patient management, and d) In women 30 years and older, the cobas HPV Test can be used to assess the presence or absence of HPV genotypes 16 and 18. This information, together with the physician’s assessment of cytology history, other risk factors, and professional guidelines, may be used to guide patient management.” However, the results from the ATHENA study and others (Castle et al., 2011; Gage, Sadorra, Lamere et al, 2012; Preisler, Rebolj, Untermann et al, 2013; Stoler et al., 2011; Wright, Stoler, Sharma et al, 2011) have led to the proposal of a subcommittee to the FDA for replacement of the classic Pap test from the cobas HPV test.

Anyplex™ II HPV28 Detection

Anyplex™ II HPV28 Detection (Seegene, Seoul, Korea) is a Real-Time PCR test, that simultaneously detects 19 high-risk HPV types (16, 18, 26, 31, 33, 35, 39, 45, 51, 52, 23, 26, 28, 29, 66, 68, 69, 73, 82) and 9 low-risk HPV types (6, 11, 40, 42, 43, 44, 54, 61, 70) using DPO™ (Dual Priming Oligonucleotides) primers and melting curve analysis method of TOCE™ (Tagging Oligonucleotide Cleavage and Extention).

This new real-time PCR assay was compared to Hybrid Capture2 (Kwon, Roh, Park et al, 2014). The Anyplex II HPV28 assay is analytically more sensitive in the detection of the 13 HR types represented by the HC2 assay and exhibited a higher concordance with comprehensive genotyping based on the sequencing analysis (Kwon et al., 2014). Furthermore, the partial quantitation of viral load based on

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

TOCE could be used to increase specificity, since only a limited number of HPV positive NILM samples seem to have high viral loads (Kim, Kwon, Woo et al, 2013).

Abbott RealTime High HPV

The Abbott RealTime High Risk (HR) HPV assay is a qualitative Real-time PCR assay intended to detect 14 high risk HPV genotypes (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66 and 68) and to partially genotype 16, 18 from other 12 high risk genotypes. It utilizes a 5 color Real-time PCR Taqman hydrolysis probe reaction; one channel is used for the detection of HPV16, one for HPV18, one for beta-globin, one for the rest HPVs, and the last is used as passive reference.

Early studies comparing this test with HC2 displayed good concordance (Kaliterna,Lepej & Vince, 2009), while more recent ones agree with early findings with somewhat higher specificity than other DNA tests (Cuzick, Cadman, Mesher et al, 2013; Szarewski, Mesher, Cadman et al, 2012; Venturoli, Leo, Nocera et al, 2012).

Non-PCR Based Techniques

Direct hybridization of HPV to complementary probes on membranes (Southern blotting) was used in the early studies that led to the identification of HPVs and later on to the discovery that HPV integration to the human genome occurs (Choo,Pan & Han, 1987). However, such methods require vast amounts of high quality DNA, are time-consuming and have relatively low sensitivity. As a result their routine use for HPV has faded. More recent techniques use in-vitro hybridization and enzymatic chemiluminescence or fluorescence to identify HPV DNA presence. Two FDA approved diagnostic tests are Hybrid Capture 2 and Cervista®.

digene Hybrid Capture 2

Hybrid Capture 2 (Qiagen, U.S.A.) includes a crude nucleic acid extraction and hybridization of HPV DNA with RNA probes. The DNA:RNA hybrids are captured on a micro-plate coated with a propriety antibody that identifies such hybrids. A second antibody for hybrids covered with multiple alkaline phosphatase is added and antibody mediated chemiluminescence production occurs that is detected on a luminometer. The HC-II HR HPV DNA test is a sandwich capture molecular hybridization assay that uses a signal amplification detection method based on chemiluminescence. Results are considered positive when their measured relative light units/cutoff ratio are $\geq$1 and should be retested for verification of a positive result if the ratio is $\geq$1 and $\leq$2.5. Thirteen HR-HPV types (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, and 68) and five low risk types (6, 11, 42, 43, and 44) can be detected with this assay, without any typing ability. The declared limit of detection varies depending on HPV type, with a median of 1.09 pg/ml that is translated to 109,000 copies/ml, a LOD significantly higher than PCR based methods (Table 1).

Due to its early FDA approval the HC2 assay is considered as the golden standard of HPV DNA techniques, with all later kits being compared to it (Gravitt, Schiffman, Solomon et al, 2008; Terry et al., 2001). There is some criticism as to whether and to what extend cross-reactivity with low risk HPV types occurs and how the cut-off for positive samples should be set (Castle et al., 2008; Rebolj, Bonde, Njor et al, 2011). In the U.S. HC2 still remains the most widely test used, since apart from the limited equipment that is needed for the test execution, limited personnel training is required to perform the

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

test. Moreover, even though HC2 has a lower analytical LOD than PCR techniques, its clinical accuracy parameters, i.e. how well a positive result is reflected to a higher grade lesion, is found to be higher than PCR in some studies (Peyton, Schiffman, Lörincz et al, 1998).

Cervista®

Cervista® utilizes a proprietary technology (Invader®, Hologic, U.S.A.) where two probes are cleaved upon detection of HPV DNA and fluorescence is produced and quantified. This is performed in an isothermal reaction at 63° C by an enzyme (Cleavase®, Hologic, U.S.A.). The first probe is partially complementary to the HPV DNA, while the rest is complementary to a fluorochrome bearing hairpin oligonucleotide. Invader® uses the partial homology created by the mismatching in order for the enzyme to work. Cleavage of the fluorochrome increases produced fluorescence that is measured after 4 hours and is proposed to amplify the signal by 1-10,000,000 times. There are two versions of the test, one for detection of high risk types with no typing (16,18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68) and one for typing of HPVs 16 and18 (Johnson, Starkey, Palmer et al, 2008). The technique is a qualitative method with an internal control for human DNA and a LOD between 2500 to 10000 copies/reaction (Table 1). In a recent study Cervista displayed similar sensitivity to PCR followed by sequencing for HPVs 16 and 18 (Bartholomew, Luff, Quigley et al, 2011).

In-situ hybridization

In situ hybridization (ISH) by chromogenic or fluorescence techniques is based on the complementary pairing of a labeled probe to HPV DNA or mRNA in cervical smears of FFPE tissues. After the hybridization a signal amplification step is performed with either biotin-streptavidin, or biotinyl tyramide, or a secondary antibody with a polymer chain. This method is supported to demonstrate the precise localization of viral genomes in individual cells and is evaluated by the presence of punctuate signals for integrated virus and diffuse signals for episomal virus (Evans & Cooper, 2004). The INFORM® HPV assay (Ventana Medical Systems, Inc., AZ, USA) is based upon chromogenic ISH and includes a low-risk (6, 11, 42, 43 and 44) and a high-risk cocktail (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59 and 66). This assay has been technically but insufficiently clinically validated (Hesselink, van den Brule, Brink et al, 2004; Qureshi, Bolick, Ringer et al, 2005). Other versions include the Genpoint HPV™ (Dako, Denmark) and the Bond HPV ISH high risk cocktail (16, 18, 31, 33, 51) and low risk cocktail (6 and 11) (Leica Microsystems). Even though some studies have shown that ISH may have sensitivity similar to consensus PCR (Kelesidis, Aish, Steller et al, 2011) and morphological evaluation is able to be performed on the test slide, ISH is considered to be laborious and lacks sufficient clinical sensitivity to be used in screening.

Immunocytochemistry - Immunohistochemistry

The idea of an antibody that would be able to detect HPVs is appealing to many physicians, since they could verify their morphologic diagnosis with extra evidence; however this has been rendered from difficult to impossible. The technical limitations that arise are a result of the virus life cycle. First, HPV only multiply in epithelial cells, where the early proteins are usually expressed in low concentrations, while the late capsid proteins are only expressed in productive infections that are only a fraction of the HPV infection cycle. Finally, no typing would be available and so far there has been a lack of sensitive and

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

specific antibodies of high quality against the viral proteins. A noteworthy attempt is from a company in Germany (ViroScreen) that has produced an alkaline phosphatase (AP) immunocytochemistry and immunohistochemistry kit that is intended to be used as a prognostic marker for lesion regression. The idea is based on the fact that viral L1 protein can produced immunologic response aiding in clearance of the infection. Also, since after HPV integration L1 production is usually abolished, L1 protein positivity suggests that integration has not happened (Choi, Kang, Kim et al, 2010).

HPV mRNA Detection

Even though HPV prevalence is higher than 25% in women from developed countries the cumulative risk for cervical cancer development is only 0.9% (Dunne et al., 2007; Jemal et al., 2011). In order for HPV to be considered oncogenic the products of its E6 and E7 oncogenes have to be expressed (Scheffner & Whitaker, 2003). As a result, commercially available mRNA detection kits have been produced targeting the genes' transcripts using as starting material either liquid based cytology samples or fresh-frozen samples, due to mRNA's rapid degradation. More recently a commercially available kit was designed for FFPE tissues using ISH for HPV mRNA detection (RNAscope, ACD Inc, U.S.A.) (Ukpo, Flanagan, Ma et al, 2011).

NASBA

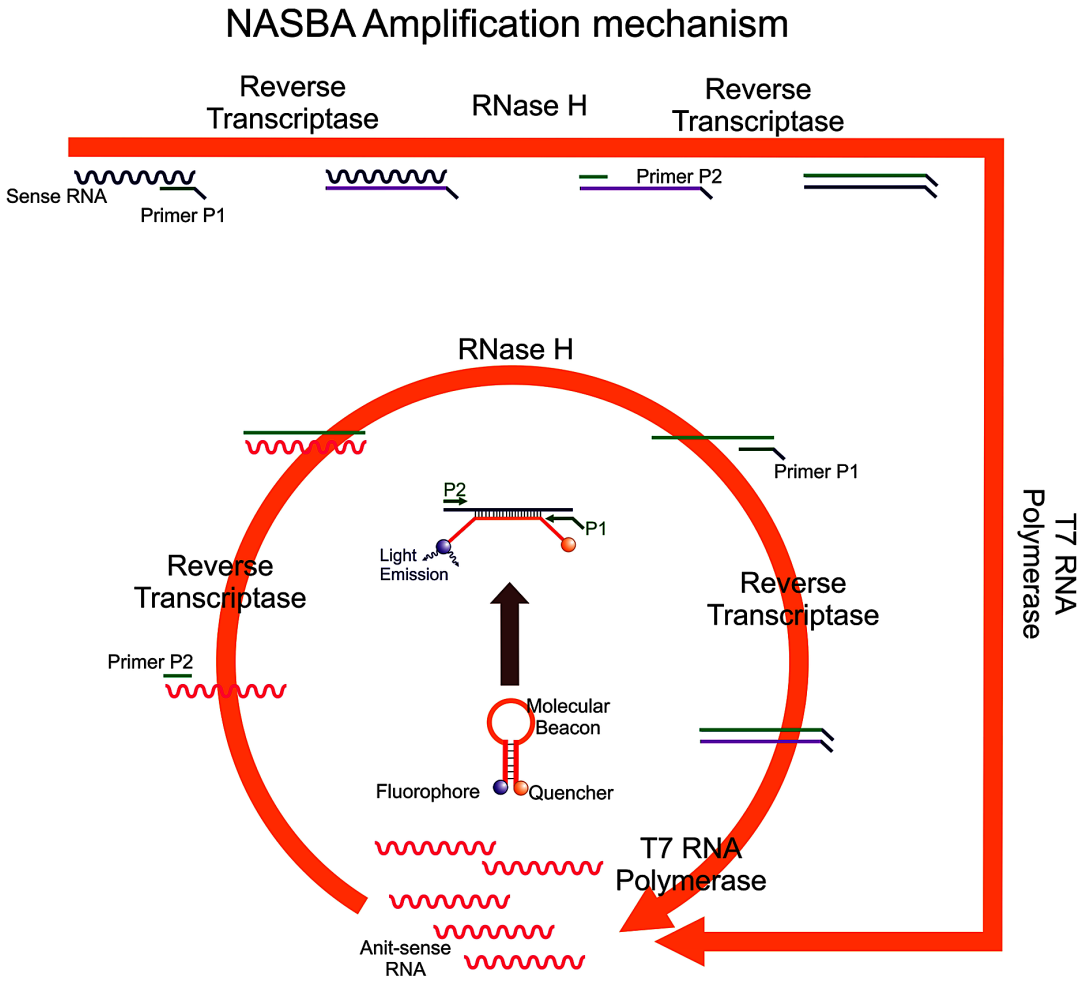
Nucleic Acid Sequence Based Amplification is a technique that uses an isothermic reaction to amplifying RNA using a RNA polymerase (Figure 4). The method uses two short single stranded DNA primers, P1 and P2 and a total of three enzymes. P1 is partially complementary to the sequence to be amplified and the rest contains a T7 RNA polymerase promoter site while P2 is a common primer. After the binding of P1 or P2 to the RNA reverse transcription occurs by AMV and the original RNA of the cDNA:RNA hybrid is hydrolysed by a RNAse H. After the binding of the second primer a dsDNA sequence is produced by AMV that had the RNA polymerase binding site. As a result, the RNA polymerase creates multiple mRNAs that are further amplified. Amplification is monitored via fluorescence production from the hybridization of molecular beacons to the products and detection in a Real-time PCR variation platform. Lack of any denaturation step inhibits any amplification from DNA sequences of the virus present. A commercial IVD version (Nuclisens EasyQ® HPV, Biomerieux, France) utilizes primers for E6 and E7 viral oncogene transcripts and type specific molecular beacons for 5 high risk HPV types (16, 18, 31, 33 and 45) and an internal control (UA1) (Smits, van Gemen, Schukkink et al, 1995). The test uses total nucleic acid extracts using the BOOM method from ThiPrep® stored cervical samples and offers type identification of HPV16, HPV18, HPV31, HPV33 and HPV45 (Molden, 2005; Molden, 2006; Sørbye, 2010). Nuclisens EasyQ® HPV has been shown to be more specific than DNA test, more effective in identifying HSILs after treatment than repeat cytology and more accurate in identifying women with HSILs reducing revisits and referral colposcopies (Molden, Nygård, Kraus et al, 2005; Sorbye, Arbyn, Fismen et al, 2011; Sorbye, Fismen, Gutteberg et al, 2010) (Table 1).

Aptima HPV Assay

A similar amplification method is used in a second commercial test (APTIMA HPV Assay, Gen-Probe, U.S.A.) that targets E6/E7 mRNA from 14 high risk HPV genotypes (16, 18, 31, 33, 35, 39, 45, 51,

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 4. The Nucleic Acid Sequence Based Amplification (NASBA) mechanism. RNAs are depicted as frizzled while DNAs as straight lines. The initial step produces a DNA sequence with an introduced T7 promoter region resulting in creation of multiple anti-sense RNA sequences. In a second cycle these RNAs reproduce the DNA sequence with the T7 promoter. All reactions happen simultaneously in a steady temperature,



52, 56, 58, 59, 66 and 68). Again the original sequence is amplified with the induction of a RNA polymerase promoter sequence in one of the primers and a second Reverse transcription enzyme (MMLV) is responsible for the DNA:RNA hybrid creation and hydrolysis (Transcription-Mediated Amplification, TMA). Detection of produced amplicons is based on the Hybridization Protection Assay in which, after the addition of complementary chemiluminescent labeled nucleic acid probes, light is emitted from the labeled RNA:DNA hybrids and is detected in a luminometer. Results are reported from the software that calculates signal to cut-off ratios from the controls and the samples tested. A major advantage for large scale testing laboratories is the fully automated versions of the test in analytes like PANTHER and TIGRIS DTS systems. ThiPrep® stored cervical samples are used as a starting material after mRNA extraction and isolation (Table 1). A prototype of this assay was evaluated in 536 women with histologi-

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

cal outcomes (Castle, Dockter, Giachetti et al, 2007). Detection of E6/E7 mRNA was strongly correlated with severity of the lesion; more than 90% of CIN 3 cases and all five carcinomas revealed E6/E7 mRNA. However, in a second study Aptima displayed poor specificity (Szarewski et al., 2008). More recent studies support that Aptima had similar sensitivity to HC2 with improved specificity (Arbyn, Roelens, Cuschieri et al, 2013).

mRNA FLOW FISH

More recently fluorescent probes for intracellular hybridization with E6 and E7 viral mRNA has been combined with flow cytometry in order to be able to quantify the percentage of cells that over-produce viral mRNA (Narimatsu & Patterson, 2005). A commercial IVD kit has been developed (HPV Oncotect, InCellDx, U.S.A.) that is currently under FDA review for approval. The test uses cells from LBC that are washed, fixed with a propriety buffer and hybridized with a cocktail of probes designed complementary to E6/E7 gene mRNAs. No denaturation is performed, so dsDNA that could be a source of false positives cannot be hybridized to the probe. The sample is analyzed on a flow cytometer with a common blue laser (488nm) and according to a cut-off of 2%, set from cytological negative samples, percentage of FAM fluorescent cells is measured and the sample is characterized as positive or negative (Table 1).

This kit has been compared to HPV DNA testing in two studies, one versus Hybrid Capture 2 and one versus CLART2 arrays, where authors determined that "with similar sensitivity and greater specificity/positive predictive value, HPV E6, E7 mRNA quantification in cells is an improvement over HPV DNA for cervical cancer screening" (Coquillard et al., 2011; Kottaridi, Tsiodras, Spathis et al, 2011; Spathis et al., 2012) Furthermore, flow cytometry allows discrimination and quantification of the cellular populations present in the cervical sample that in turn allows the characterization of sample adequacy (Kottaridi et al., 2010; Polina, Sturgis, Patterson et al, 2008). Should, in the future, a combination of HPV mRNA testing with monoclonal antibodies for the different cell populations be created, it would even be possible to quantify viral load in different epithelial cells (i.e. endocervical or ectocervical, different maturation stages of ectocervical cells and so on), allowing the monitoring of the course of viral infection.

Combinatorial Techniques, Towards Personalised Medicine

Since molecular HPV detection techniques were established, there are numerous published studies that have attempted to clarify the role of each technique as a unique test to substitute the classical test Papanicolaou (Carozzi, 2007; Coquillard et al., 2011; Cuschieri & Wentzensen, 2008; Cuzick, Arbyn, Sankaranarayanan et al, 2008; Denton, Bergeron, Klement et al, 2010; Kottaridi et al., 2011; Mathew & George, 2009; Mayrand, Duarte-Franco, Rodrigues et al, 2007; Narimatsu & Patterson, 2005; Naucler, Ryd, Tornberg et al, 2009; Smits et al., 1995; Sorbye et al., 2010; Trope, Sjoborg, Eskild et al, 2009; Tsoumpou, Arbyn, Kyrgiou et al, 2009). However there is no single molecular biology method that can reliably determine the risk for each woman to have cervical intraepithelial neoplasia. In contrast, evidence supports that the sensitivity of test Papanicolaou combined with one or more HPV DNA tests is higher than the sensitivity of each single individual method (Cuzick et al., 2008; Karakitsos, Pouliakis, Meristoudis et al, 2011; Mathew & George, 2009; Mayrand et al., 2007; Naucler et al., 2009; Spathis et al., 2012).

Computer science and especially artificial intelligence have paved the way for the development of computerized systems supporting prognosis, diagnosis and therapeutic or treatment decisions. There

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

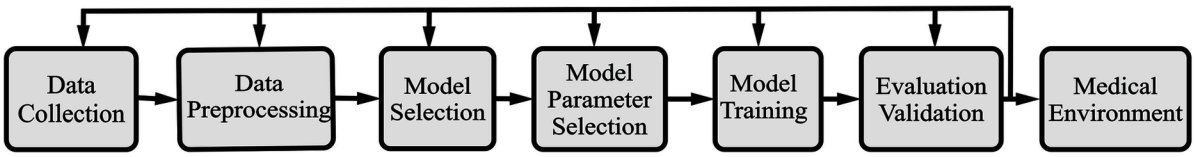
is a plethora of methods already applied, such as: neural networks (Astion & Wilding, 1992; Cochand-Priollet, Koutroumbas, Megalopoulou et al, 2006; Karakitsos, Chrelias, Pouliakis et al, 2012; Karakitsos, Cochand-Priollet, Guillausseau et al, 1996; Karakitsos, Stergiou, Pouliakis et al, 1996; Marchevsky,Tsou & Laird-Offringa, 2004; Markopoulos, Karakitsos, Botsoli-Stergiou et al, 1997; Pantazopoulos, Karakitsos, Iokim-Liossi et al, 1998), discriminant analysis (Karakitsos, Megalopoulou, Pouliakis et al, 2004; Marchevsky et al., 2004; McHugh,Shinn & Kay, 2000), classification and regression trees (CARTs) (Breiman, Friedman, Stone et al, 1984; Karakitsos et al., 2011; Pouliakis, Margari, Margari et al, 2013; Sá, 2007; Tzivras, Megalopoulou, Pouliakis et al, 2008) and genetic algorithms (Saiti, Naini, Shoorehdeli et al, 2009) among others.

In this section, an analysis related to the combinatorial use of HPV test results and classical methods, such as the test Papanicolaou (in Liquid Based Cytology (LBC) smears) is presented. The goal is to improve performance and obtain a more accurate decision for each individual woman. This involves the construction of computerized supporting medical decisions system. A typical cycle for the creation of a useful system (Figure 5) involves several steps.

The first step is data collection; the size of data is strongly related to the nature of the problem the model is called to solve, thus data size may be augmented after evaluation of the model's performance. In this step feature extraction and selection algorithms are applied, in order to create a dataset appropriate to the model and problem solution requirements. The second step is related to data preprocessing, such as removal of inconsistent entries, application of mathematical transformations to map alphanumeric data to numbers suitable for the specific model or normalization. In this step the dataset is separated into at least two sets: the training set, used for the "training" of the model, and the test set, for model performance evaluation and model validation. The third step involves the selection of a model suitable for the classification problem as well as the definition of model parameters (step four). In the sequel, follows the model "training", i.e. the application of the training set data and the model creation. During this step there are applied sophisticated algorithms. The target is to maximize the system performance in the training data, however without loss of accuracy on the test set data. The last step is the evaluation of the model performance on the test dataset, or some times on a separate part of data (validation set). If there are obtained good results in this step, then the model can be used in a production environment for routine or experimental use. Otherwise, all steps should are evaluated, and the process may be start again going one or more steps backwards, until satisfactory performance is achieved.

A preliminary study was presented in 2011 (Karakitsos et al., 2011). The study was based on 808 cases with results from cytological examination, expressed according to the Bethesda classification system (TBS2001 system) (Henry, 2003; Smith, 2002), HPV DNA typing of 35 genotypes (CLART® HPV2 test (Genomica), (Gomez-Roman, Echevarria, Salas et al, 2009), HPV mRNA typing of HPVs 16, 18, 31, 33 and 45 (NASBA, NucliSENS EasyQ® HPV v1.0, Biomerieux) (Tyagi,Bratu & Kramer, 1998), HPV High risk mRNA detection (HPV OncoTect™ E6, E7 mRNA Kit, IncellDx) (Narimatsu &

Figure 5. Typical cycle for the construction of a computerized system supporting medical decisions.



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Patterson, 2005) and immunocytochemical expression of p16 (CINtec® Cytology Kit) (Klaes, Benner, Friedrich et al, 2002). The authors tried to discriminate CIN 2 or worst cases in histology from CIN1 or negative cases. From the 808 women 399 were clinically negative (thus no histological confirmation was available) while the remaining were histologically confirmed. The study team applied a CART model to classify the cases, as well as, to create a classification algorithm. According to these preliminary results the sensitivity, specificity, PPV, NPV and overall accuracy on the combined training and test sets was: 85.12%, 92.58%, 66.88%, 97.25% and 91.46% respectively. The classification algorithm (Karakitsos et al., 2011), was relatively simple having as first triage criterion the cytological outcome and subsequently mRNA of HPV 16 and DNA of HPV 16; all other variables were excluded from the classification algorithm. These results provided evidence that more detailed studies are required.

In a second attempt to combine the outcome of the Papanicolaou test with the results of HPV related molecular ancillary examinations (Karakitsos et al., 2012), the same team employed the Learning Vector Quantizer (LVQ) artificial neural network (ANN). The data set was composed of 1258 samples being either histologically confirmed or clinically negative. In that data-set 155 cases (12.3%) had CIN2 or worse, while 1103 (87.7%) had CIN1 or normal outcome. The combined results of the LVQ, both for the training and test sets, displayed sensitivity rate of 85.16%, specificity 98.01%, PPV 85.71%, NPV 97.92% and overall accuracy of 96.42%. The results were in favor the LVQ NN in comparison to cytology alone, as the comparison of the ROC curve for cytology (A.U.C.=0.866 S.E.=0.016) and for the LVQ (A.U.C.=0.916 S.E.=0.017) proved that the LVQ ANN classifier performance was statistically better compared to the cytological diagnosis alone ($z=-2.142$, $p<0.05$). The LVQ ANN model identified more accurately women that required immediate referral to colposcopy and also reduced the number of women with clinically insignificant lesions that would be referred for colposcopy using current triaging routines. According to these results, for the LVQ ANN, 15% women required unnecessary colposcopy, whilst only 2% with a potentially significant lesion failed to be referred appropriately for further colposcopic assessment, in contrast for the cytological examination alone. Using as cut-off value ASCUS+ for referral to colposcopy the rates were 63% and 0.5%, respectively, and using ASCH+ as cut-off the percentages were 20% and 4% respectively. Therefore the combinatorial approach gave more accurate and balanced results.

In both efforts (Karakitsos et al., 2012; Karakitsos et al., 2011) analyzed previously, there was discrimination between two categories; CIN 2+ (includes CIN 2, CIN 3 and carcinomas) and CIN 1- (includes CIN 1, HPV and negative cases). In a more recent study discrimination among four groups; according to the final outcome (histological examination outcome or a clinically negative result if histological results were not available) was performed (Bountris, Haritou, Pouliakis et al). The study involved 740 cases and the classification was into four classes: Class-1: negative or clinically negative, Class-2: CIN1, Class-3: CIN2 or CIN3 and Class-4: Squamous Cell Carcinoma (SCC) or adenocarcinoma. In this study two ANNs were used in order to construct a Clinical Decision Support System (CDSS). The first ANN was a Probabilistic Neural Network (PNN), based on Bayesian classification (Specht, 1990; Wasserman, 1993) and responsible to classify a case only if the cytological outcome was not ASCUS. The second ANN was a multilayer perceptron (MLP) trained with the back propagation of the error algorithm (Duda,Hart & Stork, 2001) and responsible to classify the cytologically ASCUS cases. Obviously, via this approach the fist discriminating characteristic was the cytological examination, as in the previously mentioned two studies. In this study data were separated into three sets (training set: about 60%, test set: about 20% and validation set: about 20%). The PNN part of the CDSS had an overall accuracy (on all data as a combined set) equal to 87.7% and had satisfactory results for the discrimination among the four classes,

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

the MLP part of the CDSS allowed correct discrimination of the cytology ASCUS cases in the range of 80%, giving excellent results for this ambiguous diagnostic category. Finally the proposed CDSS was compared with other systems (based on alternative classification technologies) and was proved superior.

Issues, Controversies, Problems

Cervical cancer still remains a public health issue ranking 3rd in developing countries and 9th in developed ones, even though prophylactic vaccination for HPV is expected to reduce its incidence (Jemal et al., 2011). Still, even vaccination age, efficacy and target population is variable in different countries, due to different epidemiology of the virus (Muñoz, Bosch, Castellsagué et al, 2004). Furthermore, since most viral infections are transient in younger women (<30 years old) and HPV prevalence is highest in that age group, official guidelines in the U.S. suggest DNA testing only for either borderline cytology results, or for women over 30 or after treatment of intraepithelial lesions (Wright et al., 2007). There are several debates that have raised in the use of HPV detection techniques in screening and follow up of women that have undergone therapeutic intervention, and also for the effects of that therapeutic intervention. Some have been discussed in the introduction and some will be analyzed later on.

HPV DNA Typing

In earlier studies, since HPV 16 and 18 were the first to be recognized as the high risk types connected to cancer development, many studies identified one or both of them in over 90% of carcinomas of the cervix (Pirog, Kleter, Olgac et al, 2000). Later on after a larger poll of epidemiological data were available, that percentage dropped to 65-70% (Munoz et al., 2003). Such a result could be caused by the close phylogenetic relation of most common high risks and some types as 33 and 31 could be initially be identified as 16. Many companies have based their strategies in HPV detection products on the epidemiologic maps created by such studies; i.e. Nuclisens test type selection was probably based on the 5 most common HPV types of some studies (Muñoz et al., 2004). This may not be the exact case in all countries since variance has been observed in epidemiological data in different countries (Bosch, Manos, Muñoz et al, 1995).

Furthermore, there is still some debate on the characterization of HPV types and their epidemiological based stratification as low and high risks. Depending on the population, the number of infections available from some HPV types in each study and the method used there have been changes to the number of HPV types that are considered high risk (Dunne et al., 2007; Munoz et al., 2003). This is reflected on commercial products with a typical example being the LINEAR ARRAY HPV and CLART HPV2 where the first considers HPV26, 53 and 70 as low risk and the latter as high risk. According to HPV taxonomy and phylogenetic analysis HPVs 26, 51 and 82 belong in the same type category, stating their closer lineage to other HPV types, but one of the three is considered by some as low risk while the other two as high. This fact makes the reporting of typing results more complicated, not only in studies that use statistical analysis dichotomizing HPVs as high and low, but also in the clinical routine setting where a patient with the same type would be informed by two different tests that they are infected with a probably high risk or a probably low HPV type!

There is a growing trend in creation and implementation in screening routines of DNA tests that identify only 14 types (HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and 68) that are acknowledged as high risk types, with minimal workload for laboratories since most use Real-time PCR and automated

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

analysis. However, most CIN2+ cases have more than one HPV present and geographic prevalence of HPVs varies significantly. Limitation to the HPV types detected may reduce cost and work-load, but it would lead to underestimation of the contribution of other HPV types in CxCa development.

Last, due to the integration of HPV, that is proposed to occur in carcinomas, the L1 gene is commonly lost during integration. However, the L1 gene is the target of amplification of most commercially available tests. As a result, a percentage of 5-10% of cancers is HPV negative with L1 detection assays. In a study 69% of originally negative carcinoma cases, examined with type specific PCR for E7 gene revealed at least one high risk HPV type (Walboomers et al., 1999). This effect could cause an underestimation of the participation of some types in cervical cancer development.

Clinical and Analytical Sensitivity and Specificity

As described in each technique, every method has a different limit of detection. Techniques that only use signal amplification require a 10-100 fold higher viral load than PCR based techniques, while Real-time PCR can further lower the limit of detection to a single copy/reaction, in theory at least. However, we must not forget that analytical sensitivities that are reported by most manufacturers are accomplished using plasmids, from highly trained personnel, in an ideal environment. Cervical samples, unfortunately, may be treated with less kindness in everyday practice, with sub-optimal transportation, storage and so on. Even more, cervical samples have variable cellularities, depending on the person who collected the sample and variable percentage of leukocytes and blood that can result in false-negatives and inhibition of the test. Also the choice of DNA extraction (manual vs. automated) can have an effect in HPV detection (Chranioti, Aga, Margari et al, 2011).

Furthermore, the user that executes the technique has a profound effect on the result. A typical example is reflected from the WHO HPV Typing Proficiency test (Eklund,Zhou & Dillner, 2010). 43 samples, containing plasmids from 16 HPV types (6, 11, 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66 and 68) at various concentrations diluted in human placental DNA, were sent to 54 laboratories worldwide. Proficiency was established when at least 50 IU of HPV16 or HPV18 DNA and of 500 genome equivalents (GE) of the other 14 HPV types was established. Out of the 84 datasets gathered using 21 different assays only 71.5% had a proficiency of over 80%. The major oncogenic HPV types (16 and 18) were detected in 89.7% and 92.2% of the data sets, respectively, while HPV types 56, 59, and 68 were the least commonly detected types ($\leq$80%). This study has two significant findings; first that even from a defined sample different techniques and users display significant variance in result extraction and second that "circulating HPV typing tests might systematically underestimate the prevalence of HPV52, -56, and -59 compared to HPV16 and -18" (Eklund et al., 2010). In the second round of the HPV proficiency test results were better, but similar discordances were found (Eklund et al., 2014). Even when the user is the same, different kits exhibit some variance depending on the technique and setting used (Cuzick et al., 2013; Rebolj et al., 2014; Szarewski et al., 2012)

As described earlier, due to different LODs detection rates, positivity between PCR based techniques and signal amplification techniques as Cervista and HC2, differs. Most studies performed have reached the conclusion that sensitivity of signal amplification techniques is adequate to tackle the viral load present in higher grade lesions. Super-sensitive techniques, as are typing ones, may have higher HPV detection rates, but that results in a penalty in specificity for high grade lesion detection, since even cytological and histological negative samples may carry a transient infection from a high risk HPV type (Gravitt et al., 2008; Tsiodras et al., 2011). In this setting a test with lower analytical sensitivity can have similar

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

"clinical" sensitivity; i.e. it can detect the women in risk with similar adequacy as a highly-sensitive test. Judging through the eye of a health professional, one would always prefer the most sensitive technique and the one providing most available data for analysis, with higher detection rates and reporting of low risk types. On the other hand a patient would probably prefer a test with the greatest specificity; i.e. she would be interested to know only if she is at risk of cancer development and not if she is infected with a low risk HPV type.

DNA vs. RNA Techniques

HPV DNA detection is currently the only ancillary technique that has been included in guidelines for patient management in screening and triaging women (Cuzick et al., 2008; Wright et al., 2007). Technically DNA is more stable as a molecule, easier to handle and to amplify that RNA, while there are more tests available to choose from. Furthermore, DNA testing is also able to provide typing of the virus that has offered substantially in the creation of epidemiology maps in each country and vaccination strategies for each country, since HPV type prevalence is different. Also, typing of the virus can provide valuable information of persistent infection from a specific type and information as to the combination of viruses found in multiple infections. So far, there are few data on the effect of variable combinations of high risk types in the development of lesions, or if there is any preference in combinations of HPVs (Rousseau, Pereira, Prado et al, 2001). Finally, typing will play an important role in monitoring of changes in type prevalence of HPV in vaccinated populations.

On the other hand, mRNA based techniques are able to provide information about disease progression, since they detect true oncogenic potential of the HPV infection. mRNA techniques may suffer in sensitivity, mainly because they are able to detect only a portion of all high risk HPV types, but have significantly higher specificity values in screening populations in the identification of high grade lesions (Cuzick et al., 2013; Rebolj et al., 2014; Sorbye et al., 2010; Spathis et al., 2012; Szarewski et al., 2012). However, few commercially available tests and few large cross-sectional studies with long term follow-up, in order to verify their potential, have been performed. Nevertheless, mRNA testing is probably the future in HPV detection and monitoring.

FUTURE RESEARCH DIRECTIONS

There are many open subjects still in HPV research and in their implications in future screening and triaging strategies. The first major axis relies on the evolution of current techniques use and the creation of new tests that will be able to identify women at true risk for cancer development. Up until now much effort has been targeted in creating automated techniques for rapid and large scale HPV DNA detection that has provided significant information in HPV epidemiology and its prevalence in various grades of lesions and geographic areas. However it is clear that L1 DNA detection has limitations both due to HPV integration and the fact that most HPV infections, even from high risk types, are transient and of little clinical value. Future HPV typing based on the E6 or E7 gene would abolish part of the project, while typing using mRNA as target would resolve these drawbacks. An automated Real-time PCR based RT-PCR would be the ideal way to go. This would also provide definite quantitative data on the viral infection course. The multiparameteric flow cytometric evaluation of HPV mRNA in all the distinct populations present in a cervical sample would also be of great interest. Finally, even though we have

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

not made any reference in this chapter, since it is oriented in HPV detection, there is a growing trend for analysis of epigenetic alterations during cervical lesion development.

The second major axis, at which HPV research should be targeted, for us, is data analysis and interpretation evolution. It is clear from many other instances that in order to approach a complex problem, such as cancer, a single test or approach is usually futile. Even though cervical cancer is directly linked to HPVs and the major risk factor is already known, CxCa still remains a complex problem. There are many techniques from which many data are available, but usually these have been limited in pair wise comparison of two or three techniques, validating their use in clinical practice. Maybe it is time for a more holistic approach. A large cross-sectional study with long term follow-up that would include DNA, RNA, protein, integration status, viral load and epigenetic alterations would probably provide the necessary data in order to create a model for true risk assessment and creation of screening and triaging strategies. Some pivotal studies that utilize complex artificial networks for diagnosis optimization and risk assessment have been described previously (Combinatorial techniques, towards personalised medicine). Even though, the uses of LVQ and CDSS have shown a significant increase in diagnostic accuracy and triaging efficiency, the scale of these studies limits their effects. A large scale study would provide significant data as to selection of reflex test and triaging options with the highest accuracy and the lowest cost. Thus a new

For either of these axes to be able to succeed there are many prerequisites. The cataclysm and diversity of data that would be gathered from such a large scale study would have to be collected, digitized and managed centrally by a certified organization. Unlike a cancer registry where only data from patients suffering from the disease are gathered, the HPV registry would have to be way more complex, as it should contain all medical examinations and patient's history, throughout a period of years to decades. Apart from the obvious problem of which that organization that would have the vast amount of data should be, problems arise also in the security of these files. Since a large scale monitoring attempt would require users from distant regions to enter data, a web based application is essential. However, this creates risks and requires encrypted data-bases to be created for securing these medical e-data. Furthermore, selection of users and their access privileges raise serious issues, since apart from the patient's physician who will have to have access their patients' e-health file, some other users must have access in order for these data to be analyzed. This analysis, that can only be performed with bioinformatics, is essential for the translation of the data in a form that can be used for classification and prediction modeling, for validation of efficiency and accuracy of new molecular tests and for the creation of new clinical decision systems that can provide novel clinical guidelines and ultimately a true personalized based risk assessment and triaging system. An issue in this effort is the interpretation of decision systems, some of them such as the CART produce human understandable algorithms (some time very simple) that may be used in the everyday practice and to knowledge extraction and discovery, however other methods (for example the LVQ based methods or the back propagation neural networks) are so complex that is nearly impossible (if not impossible) for the average user to understand the classification process and rules, thus such models are usable only via the use of a computer, hopefully modern technology can provide access to these models even through mobile phones.

Preliminary results suggested that the incorporation of new tests and combinations of biomarkers may improve the accuracy of diagnosis, thus reducing overload of colposcopy clinics and guide to personalised management and intervention. Apart from the need, as already mentioned, to confirm the reproducibility of such results in large scale datasets, it is also obvious that molecular biology tests increase cost; different combinations with a selection of the most appropriate tests may prove to have

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

similar accuracy with reduced cost. Further studies assessing also the cost and financial implication of such an approach on screening programs are required and detailed cost-effectiveness analyses are still required and should continuously evolve.

CONCLUSION

Since the initial identification of HPVs and their linking to cancer development, much has changed. We now have a vast arson of high quality and throughput techniques available for HPV detection, monitoring and identification of molecular changes that happen in cervical cells. Each technique has its' pros and cons, while choice of a single test provides some info for the clinician, but little info about the outcome of the patient. Integration of multiple methods in a large scale populations would allow vast amount of data that to be gathered, requiring the creation of a complex information storage and management computer system that, after training, should be able to identify, early and accurately, women in true risk for cervical cancer development, leading to a true personalized patient management.

REFERENCES

Andersson, S., Safari, H., Mints, M., Lewensohn-Fuchs, I., Gyllensten, U., & Johansson, B. (2005). Type distribution, viral load and integration status of high-risk human papillomaviruses in pre-stages of cervical cancer (CIN). *British Journal of Cancer*, *92*(12), 2195–2200. doi:10.1038/sj.bjc.6602648 PMID:15942630

Arbyn, M., Roelens, J., Cuschieri, K., Cuzick, J., Szarewski, A., Ratnam, S., Reuschenbach, M., Belinson, S., Belinson, J. L., & Monsonego, J. (2013). The APTIMA HPV assay versus the Hybrid Capture 2 test in triage of women with ASC-US or LSIL cervical cytology: a meta-analysis of the diagnostic accuracy. *International Journal of Cancer, 132*(1), 101-108.

Astion, M. L., & Wilding, P. (1992). Application of neural networks to the interpretation of laboratory data in cancer diagnosis. *Clinical Chemistry*, *38*(1), 34–38. PMID:1733603

Bartholomew, D. A., Luff, R. D., Quigley, N. B., Curtis, M., & Olson, M. C. (2011). Analytical performance of Cervista HPV 16/18 genotyping test for cervical cytology samples. *Journal of Clinical Virology*, *51*(1), 38–43. doi:10.1016/j.jcv.2011.01.016 PMID:21376660

Bartlett, J. M. S., & Stirling, D. (2003). A short history of the polymerase chain reaction. *Methods in Molecular Biology, 226*(3-6.

Bosch, F. X., Manos, M. M., Muñoz, N., Sherman, M., Jansen, A. M., Peto, J., & Shah, K. V. et al. (1995). Prevalence of human papillomavirus in cervical cancer: A worldwide perspective. International biological study on cervical cancer (IBSCC) Study Group. *Journal of the National Cancer Institute*, *87*(11), 796–802. doi:10.1093/jnci/87.11.796 PMID:7791229

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Bountris, P., Haritou, M., Pouliakis, A., Margari, N., Kyrgiou, M., Spathis, A., & Koutsouris, D.-D. et al. (2014). An Intelligent Clinical Decision Support System for Patient-Specific Predictions to Improve Cervical Intraepithelial Neoplasia Detection. *BioMed Research International*, *2014*, 341483. doi:10.1155/2014/341483 PMID:24812614

Bouvard, V., Baan, R., Straif, K., Grosse, Y., Secretan, B., El Ghissassi, F., & Cogliano, V. et al. (2009). A review of human carcinogens--Part B: Biological agents. *The Lancet Oncology*, *10*(4), 321–322. PMID:19350698

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. Belmont, California: Wadsworth International Group.

Carozzi, F. M. (2007). Combined analysis of HPV DNA and p16INK4a expression to predict prognosis in ASCUS and LSIL pap smears. *Coll Antropol, 31*(2), 103-106.

Castle, P. E., Dockter, J., Giachetti, C., Garcia, F. A., McCormick, M. K., Mitchell, A. L., & Kolk, D. P. et al. (2007). A cross-sectional study of a prototype carcinogenic human papillomavirus E6/E7 messenger RNA assay for detection of cervical precancer and cancer. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, *13*(9), 2599–2605. doi:10.1158/1078-0432.CCR-06-2881 PMID:17473189

Castle, P. E., Solomon, D., Wheeler, C. M., Gravitt, P. E., Wacholder, S., & Schiffman, M. (2008). Human papillomavirus genotype specificity of hybrid capture 2. *Journal of Clinical Microbiology*, *46*(8), 2595–2604. doi:10.1128/JCM.00824-08 PMID:18579716

Castle, P. E., Stoler, M. H., Wright, T. C., Sharma, A., Wright, T. L., & M, B. C. (2011). Performance of carcinogenic human papillomavirus (HPV) testing and HPV16 or HPV18 genotyping for cervical cancer screening of women aged 25 years and older: a subanalysis of the ATHENA study. *The Lancet Oncology, 12*, 880-890.

Choi, Y. S., Kang, W. D., Kim, S. M., Choi, Y. D., Nam, J. H., Park, C. S., & Choi, H. S. (2010). Human papillomavirus L1 capsid protein and human papillomavirus type 16 as prognostic markers in cervical intraepithelial neoplasia 1. *International Journal of Gynecological Cancer: Official Journal of the International Gynecological Cancer Society*, *20*(2), 288–293. doi:10.1111/IGC.0b013e3181cd184c PMID:20134272

Choo, K. B., Pan, C. C., & Han, S. H. (1987). Integration of human papillomavirus type 16 into cellular DNA of cervical carcinoma: Preferential deletion of the E2 gene and invariable retention of the long control region and the E6/E7 open reading frames. *Virology*, *161*(1), 259–261. doi:10.1016/0042-6822(87)90195-4 PMID:2823467

Chranioti, A., Aga, E., Margari, N., Kottaridi, C., Pappas, A., Panayiotides, I., & Karakitsos, P. (2011). Performance evaluation of manual and automated (MagNA pure) nucleic acid isolation in HPV detection and genotyping using Roche Linear Array HPV Test. *Infectious Diseases in Obstetrics and Gynecology*, *2011*(931281).

Chranioti, A., Spathis, A., Aga, E., Meristoudis, C., Pappas, A., Panayiotides, I., & Karakitsos, P. (2012). Comparison of two commercially available methods for HPV genotyping: CLART HPV2 and Linear Array HPV Genotyping tests. *Analytical and Quantitative Cytology and Histology*, *34*(5), 257–263.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Cochand-Priollet, B., Koutroumbas, K., Megalopoulou, T. M., Pouliakis, A., Sivolapenko, G., & Karakitsos, P. (2006).Discriminating benign from malignant thyroid lesions using artificial intelligence and statistical selection of morphometric features. *Oncology Reports*, *15*, 1023–1026.

Coquillard, G., Palao, B., & Patterson, B. K. (2011).Quantification of intracellular HPV E6/E7 mRNA expression increases the specificity and positive predictive value of cervical cancer screening compared to HPV DNA. *Gynecologic Oncology*, *120*, 89–93.

Cuschieri, K., & Wentzensen, N. (2008). Human papillomavirus mRNA and p16 detection as biomarkers for the improved diagnosis of cervical neoplasia. *Cancer Epidemiology, Biomarkers & Prevention*, *17*(10), 2536–2545. doi:10.1158/1055-9965.EPI-08-0306 PMID:18842994

Cuzick, J., Arbyn, M., Sankaranarayanan, R., Tsu, V., Ronco, G., Mayrand, M. H., & Meijer, C. J. et al. (2008).Overview of human papillomavirus-based and other novel options for cervical cancer screening in developed and developing countries. *Vaccine*, (10): 29–41.

Cuzick, J., Cadman, L., Mesher, D., Austin, J., Ashdown-Barr, L., Ho, L., & Szarewski, A. et al. (2013). Comparing the performance of six human papillomavirus tests in a screening population.[PubMed]. *British Journal of Cancer*, *108*(4), 908–913. doi:10.1038/bjc.2013.22

Dal Bello, B., Spinillo, A., Alberizzi, P., Cesari, S., Gardella, B., & Silini, E. M. (2009). Validation of the SPF10 LiPA human papillomavirus typing assay using formalin-fixed paraffin-embedded cervical biopsy samples. *Journal of Clinical Microbiology*, *47*(7), 2175–2180. doi:10.1128/JCM.00286-09 PMID:19474270

de Villiers, E. M., Fauquet, C., Broker, T. R., Bernard, H. U., & zur Hausen, H. (2004). Classification of papillomaviruses. *Virology*, *324*(1), 17–27. doi:10.1016/j.virol.2004.03.033 PMID:15183049

Denton, K. J., Bergeron, C., Klement, P., Trunk, M. J., Keller, T., & Ridder, R. (2010). The sensitivity and specificity of p16(INK4a) cytology vs HPV testing for detecting high-grade cervical disease in the triage of ASC-US and LSIL pap cytology results. *American Journal of Clinical Pathology*, *134*(1), 12–21. doi:10.1309/AJCP3CD9YKYFJDQL PMID:20551261

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification*. New York: Wiley.

Dunne, E. F., Unger, E. R., Sternberg, M., McQuillan, G., Swan, D. C., Patel, S. S., & Markowitz, L. E. (2007). Prevalence of HPV infection among females in the United States. *Journal of the American Medical Association*, *297*(8), 813–819. doi:10.1001/jama.297.8.813 PMID:17327523

Eklund, C., Forslund, O., Wallin, K. L., Dillner, J., & Loeffelholz, M. J. (2014). Global improvement in genotyping of human papillomavirus DNA: The 2011 HPV LabNet International Proficiency Study. *Journal of Clinical Microbiology*, *52*(2), 449–459. doi:10.1128/JCM.02453-13 PMID:24478473

Eklund, C., Zhou, T., & Dillner, J. (2010). Global proficiency study of human papillomavirus genotyping. *Journal of Clinical Microbiology*, *48*(11), 4147–4155. doi:10.1128/JCM.00918-10 PMID:20844222

Evans, M. F., & Cooper, K. (2004). Human papillomavirus integration: Detection by in situ hybridization and potential clinical application. *The Journal of Pathology*, *202*(1), 1–4. doi:10.1002/path.1519 PMID:14694515

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Fontaine, V., Mascaux, C., Weyn, C., Bernis, A., Celio, N., Lefèvre, P., & Garbar, C. et al. (2007). Evaluation of combined general primer-mediated PCR sequencing and type-specific PCR strategies for determination of human papillomavirus genotypes in cervical cell specimens. *Journal of Clinical Microbiology*, *45*(3), 928–934. doi:10.1128/JCM.02098-06 PMID:17229855

Gage, J. C., Sadorra, M., Lamere, B. J., Kail, R., Aldrich, C., Kinney, W., & Castle, P. E. et al. (2012). Comparison of the cobas Human Papillomavirus (HPV) test with the hybrid capture 2 and linear array HPV DNA tests. *Journal of Clinical Microbiology*, *50*(1), 61–65. PMID:22075592

García-Sierra, N., Martró, E., Castellà, E., Llatjós, M., Tarrats, A., Bascuñana, E., & Ausina, V. et al. (2009). Evaluation of an array-based method for human papillomavirus detection and genotyping in comparison with conventional methods used in cervical cancer screening. *Journal of Clinical Microbiology*, *47*(7), 2165–2169. doi:10.1128/JCM.00402-09 PMID:19439534

Gomez-Roman, J. J., Echevarria, C., Salas, S., Gonzalez-Moran, M. A., Perez-Mies, B., Garcia-Higuera, I., & Val-Bernal, J. F. et al. (2009). A type-specific study of human papillomavirus prevalence in cervicovaginal samples in three different Spanish regions. *APMIS*, *117*(1), 22–27. doi:10.1111/j.1600-0463.2008.00009.x PMID:19161533

Gravitt, P. E., Peyton, C. L., Alessi, T. Q., Wheeler, C. M., Coutlée, F., Hildesheim, A., & Apple, R. J. et al. (2000). Improved amplification of genital human papillomaviruses. *Journal of Clinical Microbiology*, *38*(1), 357–361. PMID:10618116

Gravitt, P. E., Schiffman, M., Solomon, D., Wheeler, C. M., & Castle, P. E. (2008). A comparison of linear array and hybrid capture 2 for detection of carcinogenic human papillomavirus and cervical precancer in ASCUS-LSIL triage study. *Cancer Epidemiology, Biomarkers & Prevention*, *17*(5), 1248–1254. doi:10.1158/1055-9965.EPI-07-2904 PMID:18483347

Harper, D. M. (2009). Current prophylactic HPV vaccines and gynecologic premalignancies. *Current Opinion in Obstetrics & Gynecology*, *21*(6), 457–464. PMID:19923989

Henry, M. R. (2003). The Bethesda System 2001: An update of new terminology for gynecologic cytology. *Clinics in Laboratory Medicine*, *23*(3), 585–603. doi:10.1016/S0272-2712(03)00053-2 PMID:14560530

Herraez-Hernandez, E., Alvarez-Perez, M., Navarro-Bustos, G., Esquivias, J., Alonso, S., Aneiros-Fernandez, J., & Rodriguez-Peralto, J. L. et al. (2013). HPV Direct Flow CHIP: A new human papillomavirus genotyping method based on direct PCR from crude-cell extracts. *Journal of Virological Methods*, *193*(1), 9–17. doi:10.1016/j.jviromet.2013.04.018 PMID:23680093

Herraez-Hernandez, E., Preda, O., Alonso, S., Pardo, R. S., & Olmo, A. (2013). Detection and Genotyping of Human Papillomavirus DNA in Formalin-Fixed Paraffin-Embedded Specimens with the HPV Direct Flow CHIP System. *The Open Virology Journal, 7*, 91-95.

Hesselink, A. T., van den Brule, A. J. C., Brink, A. A. T. P., Berkhof, J., van Kemenade, F. J., Verheijen, R. H. M., & Snijders, P. J. F. (2004). Comparison of hybrid capture 2 with in situ hybridization for the detection of high-risk human papillomavirus in liquid-based cervical samples. *Cancer*, *102*(1), 11–18. doi:10.1002/cncr.11904 PMID:14968413

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Jabłońska, S., Orth, G., Jarzabek-Chorzelska, M., Rzesa, G., Obałek, S., Gliński, W., & Croissant, O. et al. (1978). Malignancies induced by human papilloma virus. *Przeglad Dermatologiczny*, *65*(5), 497–450. PMID:214825

Jemal, A., Bray, F., Center, M. M., Ferlay, J., Ward, E., & Forman, D. (2011). Global cancer statistics. *CA: a Cancer Journal for Clinicians*, *61*(2), 69–90. doi:10.3322/caac.20107 PMID:21296855

Johnson, L. R., Starkey, C. R., Palmer, J., Taylor, J., Stout, S., Holt, S., & Miller, G. C. et al. (2008). A comparison of two methods to determine the presence of high-risk HPV cervical infections. *American Journal of Clinical Pathology*, *130*(3), 401–408. doi:10.1309/4DXEAFG2JXYF34N3 PMID:18701413

Kahn, J. A. (2009). HPV vaccination for the prevention of cervical intraepithelial neoplasia. *The New England Journal of Medicine*, *361*(3), 271–278. doi:10.1056/NEJMct0806938 PMID:19605832

Kaliterna, V., Lepej, S. Z., & Vince, A. (2009). Comparison between the Abbott RealTime High Risk HPV assay and the Hybrid Capture 2 assay for detecting high-risk human papillomavirus DNA in cervical specimens. *Journal of Medical Microbiology*, *58*(Pt 12), 1662–1663. doi:10.1099/jmm.0.012138-0 PMID:19661205

Karakitsos, P., Chrelias, C., Pouliakis, A., Koliopoulos, G., Spathis, A., Kyrgiou, M., & Paraskevaidis, E. et al. (2012).Identification of women for referral to colposcopy by neural networks: A preliminary study based on LBC and molecular biomarkers. *Journal of Biomedicine & Biotechnology*, *2012*(303192).

Karakitsos, P., Cochand-Priollet, B., Guillausseau, P. J., & Pouliakis, A. (1996).Potential of the back propagation neural network in the morphologic examination of thyroid lesions.[PubMed]. *Analytical and Quantitative Cytology and Histology*, *18*(6), 494–500.

Karakitsos, P., Megalopoulou, T. M., Pouliakis, A., Tzivras, M., Archimandritis, A., & Kyroudes, A. (2004). Application of discriminant analysis and quantitative cytologic examination to gastric lesions. *Analytical and Quantitative Cytology and Histology*, *26*(6), 314–322. PMID:15678613

Karakitsos, P., Pouliakis, A., Meristoudis, C., Margari, N., Kassanos, D., Kyrgiou, M., & Paraskevaidis, E. et al. (2011). A preliminary study of the potential of tree classifiers in triage of high-grade squamous intraepithelial lesions. *Analytical and Quantitative Cytology and Histology*, *33*(3), 132–140. PMID:21980616

Karakitsos, P., Stergiou, E. B., Pouliakis, A., Tzivras, M., Archimandritis, A., Liossi, A. I., & Kyrkou, K. (1996). Potential of the back propagation neural network in the discrimination of benign from malignant gastric cells. *Analytical and Quantitative Cytology and Histology*, *18*(3), 245–250. PMID:8790840

Keegan, H., Mc Inerney, J., Pilkington, L., Gronn, P., Silva, I., Karlsen, F., & Martin, C. et al. (2009). Comparison of HPV detection technologies: Hybrid capture 2, PreTect HPV-Proofer and analysis of HPV DNA viral load in HPV16, HPV18 and HPV33 E6/E7 mRNA positive specimens. *Journal of Virological Methods*, *155*(1), 61–66. doi:10.1016/j.jviromet.2008.09.027 PMID:18955086

Kelesidis, T., Aish, L., Steller, M. A., Aish, I. S., Shen, J., Foukas, P., & Tsiodras, S. et al. (2011). Human papillomavirus (HPV) detection using in situ hybridization in histologic samples: Correlations with cytologic changes and polymerase chain reaction HPV detection. *American Journal of Clinical Pathology*, *136*(1), 119–127. doi:10.1309/AJCP03HUQYZMWATP PMID:21685039

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Kim, Y. J., Kwon, M. J., Woo, H. Y., & Paik, S. Y. (2013). Prevalence of human papillomavirus infection and genotype distribution determined by the cyclic-catcher melting temperature analysis in Korean medical checkup population. *Journal of Microbiology (Seoul, Korea)*, *51*(5), 665–670. doi:10.1007/s12275-013-3160-3 PMID:24037659

Klaes, R., Benner, A., Friedrich, T., Ridder, R., Herrington, S., Jenkins, D., & von Knebel Doeberitz, M. et al. (2002). p16INK4a immunohistochemistry improves interobserver agreement in the diagnosis of cervical intraepithelial neoplasia. *The American Journal of Surgical Pathology*, *26*(11), 1389–1399. doi:10.1097/00000478-200211000-00001 PMID:12409715

Kleter, B., van Doorn, L. J., Schrauwen, L., Molijn, A., Sastrowijoto, S., ter Schegget, J., & Quint, W. et al. (1999). Development and clinical evaluation of a highly sensitive PCR-reverse hybridization line probe assay for detection and identification of anogenital human papillomavirus. *Journal of Clinical Microbiology*, *37*(8), 2508–2517. PMID:10405393

Kottaridi, C., Georgoulakis, J., Kassanos, D., Pappas, A., Spathis, A., Margari, N., & Karakitsos, P. et al. (2010). Use of flow cytometry as a quality control device for liquid-based cervical cytology specimens. *Cytometry. Part B, Clinical Cytometry*, *78*(1), 37–40. PMID:19746413

Kottaridi, C., Tsiodras, S., Spathis, A., Chranioti, A., Pappas, A., Kassanos, D., & Karakitsos, P. et al. (2011). Clinical performance of human papillomavirus E6, E7 mRNA flow cytometric assay compared to human papillomavirus DNA typing. *Analytical and Quantitative Cytology and Histology*, *33*(6), 305–310. PMID:22590807

Kwon, M. J., Roh, K. H., Park, H., & Woo, H. Y. (2014). Comparison of the Anyplex II HPV28 assay with the Hybrid Capture 2 assay for the detection of HPV infection. *Journal of Clinical Virology*, *59*(4), 246–249. doi:10.1016/j.jcv.2014.01.015 PMID:24568964

Liu, S. S., Leung, R. C., Chan, K. K., Cheung, A. N., & Ngan, H. Y. (2010). Evaluation of a newly developed GenoArray human papillomavirus (HPV) genotyping assay and comparison with the Roche Linear Array HPV genotyping assay. *Journal of Clinical Microbiology*, *48*(3), 758–764. doi:10.1128/JCM.00989-09 PMID:20042614

Manos, M. M., Kinney, W. K., Hurley, L. B., Sherman, M. E., Shieh-Ngai, J., Kurman, R. J., & Hiatt, R. A. et al. (1999). Identifying women with cervical neoplasia: Using human papillomavirus DNA testing for equivocal Papanicolaou results. *Journal of the American Medical Association*, *281*(17), 1605–1610. doi:10.1001/jama.281.17.1605 PMID:10235153

Mantovani, F., & Banks, L. (2001). The human papillomavirus E6 protein and its contribution to malignant progression. *Oncogene*, *20*(54), 7874–7887. doi:10.1038/sj.onc.1204869 PMID:11753670

Marchevsky, A. M., Tsou, J. A., & Laird-Offringa, I. A. (2004). Classification of individual lung cancer cell lines based on DNA methylation markers: Use of linear discriminant analysis and artificial neural networks. *The Journal of Molecular Diagnostics*, *6*(1), 28–36. doi:10.1016/S1525-1578(10)60488-6 PMID:14736824

Markopoulos, C., Karakitsos, P., Botsoli-Stergiou, E., Pouliakis, A., Ioakim-Liossi, A., Kyrkou, K., & Gogas, J. (1997). Application of the learning vector quantizer to the classification of breast lesions. *Analytical and Quantitative Cytology and Histology*, *19*(5), 453–460. PMID:9349906

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Mathew, A., & George, P. S. (2009). Trends in incidence and mortality rates of squamous cell carcinoma and adenocarcinoma of cervix--worldwide. *Asian Pacific Journal of Cancer Prevention*, *10*(4), 645–650. PMID:19827887

Mayrand, M. H., Duarte-Franco, E., Rodrigues, I., Walter, S. D., Hanley, J., Ferenczy, A., & Franco, E. L. et al. (2007). Human papillomavirus DNA versus Papanicolaou screening tests for cervical cancer. *The New England Journal of Medicine*, *357*(16), 1579–1588. doi:10.1056/NEJMoa071430 PMID:17942871

McHugh, E. S., Shinn, A. P., & Kay, J. W. (2000). Discrimination of the notifiable pathogen Gyrodactylus salaris from G. thymalli (Monogenea) using statistical classifiers applied to morphometric data. *Parasitology, 121*(3), 315-323.

Molden, T., Kraus, I., Karlsen, F., Skomedal, H., & Hagmar, B. (2006). Human papillomavirus E6/E7 mRNA expression in women younger than 30 years of age. *Gynecologic Oncology, 100*, 95-100.

Molden, T., Nygård, J. F., Kraus, I., Karlsen, F., Nygård, M., Skare, G. B., Skomedal, H., Thoresen, S. O., & Hagmar, B. (2005). Predicting CIN2+ when detecting HPV mRNA and DNA by PreTect HPV-proofer and consensus PCR: A 2-year follow-up of women with ASCUS or LSIL Pap smear. *International Journal of Gynecological Cancer, 114*, 973-976.

Münger, K., Basile, J. R., Duensing, S., Eichten, A., Gonzalez, S. L., Grace, M., & Zacny, V. L. (2001). Biological activities and molecular targets of the human papillomavirus E7 oncoprotein. *Oncogene*, *20*(54), 7888–7898. doi:10.1038/sj.onc.1204860 PMID:11753671

Muñoz, N., Bosch, F. X., Castellsagué, X., Díaz, M., de Sanjose, S., Hammouda, D., Shah, K. V., & Meijer, C. J. L. M. (2004). Against which human papillomavirus types shall we vaccinate and screen? The international perspective. *International Journal of Gynecological Cancer, 111*, 278-285.

Munoz, N., Bosch, X., Sanjose, S. d., Herrero, R., Castellsague, X., Shah, K. V., Snijders, P. J. F., & Meijer, C. J. L. M. (2003). Epidemiologic Classification of Human Papillomavirus. *New England Journal of Medicine, 6*, 518-527.

Narimatsu, R., & Patterson, B. K. (2005). High-throughput cervical cancer screening using intracellular human papillomavirus E6 and E7 mRNA quantification by flow cytometry. *American Journal of Clinical Pathology*, *123*(5), 716–723. doi:10.1309/FE70AVNY75TDDJUH PMID:15981811

Naucler, P., Ryd, W., Tornberg, S., Strand, A., Wadell, G., Elfgren, K., & Dillner, J. et al. (2009). Efficacy of HPV DNA testing with cytology triage and/or repeat HPV DNA testing in primary cervical cancer screening. *Journal of the National Cancer Institute*, *101*(2), 88–99. doi:10.1093/jnci/djn444 PMID:19141778

Pantazopoulos, D., Karakitsos, P., Iokim-Liossi, A., Pouliakis, A., Botsoli-Stergiou, E., & Dimopoulos, C. (1998). Back propagation neural network in the discrimination of benign from malignant lower urinary tract lesions. *The Journal of Urology*, *159*(5), 1619–1623. doi:10.1097/00005392-199805000-00057 PMID:9554366

Peto, J., Gilham, C., Fletcher, O., & Matthews, F. E. (2004). The cervical cancer epidemic that screening has prevented in the UK. *Lancet*, *364*(9430), 249–256. doi:10.1016/S0140-6736(04)16674-9 PMID:15262102

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Peyton, C. L., Schiffman, M., Lörincz, A. T., Hunt, W. C., Mielzynska, I., Bratti, C., & Wheeler, C. M. et al. (1998). Comparison of PCR- and hybrid capture-based human papillomavirus detection systems using multiple cervical specimen collection strategies. *Journal of Clinical Microbiology*, *36*(11), 3248–3254. PMID:9774574

Pirog, E. C., Kleter, B., Olgac, S., Bobkiewicz, P., Lindeman, J., Quint, W. G., & Isacson, C. et al. (2000). Prevalence of human papillomavirus DNA in different histological subtypes of cervical adenocarcinoma. *American Journal of Pathology*, *157*(4), 1055–1062. doi:10.1016/S0002-9440(10)64619-6 PMID:11021808

Pista, A., de Oliveira, C. F., Cunha, M. J., Paixao, M. T., & Real, O. (2011). Prevalence of human papillomavirus infection in women in Portugal: the CLEOPATRE Portugal study. *International Journal of Gynecological Cancer: Official Journal of the International Gynecological Cancer Society, 21*(6), 1150-1158.

Pista, A., Oliveira, A., Verdasca, N., & Ribeiro, F. (2011). Single and multiple human papillomavirus infections in cervical abnormalities in Portuguese women. *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*, *17*(6), 941–946. doi:10.1111/j.1469-0691.2010.03387.x PMID:21040156

Pista, A., Verdasca, N., & Oliveira, A. (2011). Clinical performance of the CLART human papillomavirus 2 assay compared with the hybrid capture 2 test. *Journal of Medical Virology*, *83*(2), 272–276. doi:10.1002/jmv.21952 PMID:21181922

Polina, R., Sturgis, C., Patterson, J., & Patterson, B. K. (2008). Rapid, high throughput determination of cervical cytology specimen adequacy using a capillary-based cytometer. *Cytometry. Part B, Clinical Cytometry*, *74*(2), 133–136. doi:10.1002/cyto.b.20385 PMID:18044721

Pouliakis, A., Margari, C., Margari, N., Chrelias, C., Zygouris, D., Meristoudis, C., & Karakitsos, P. et al. (2013). (in press). Using classification and regression trees, liquid-based cytology and nuclear morphometry for the discrimination of endometrial lesions. *Diagnostic Cytopathology*. PMID:24273089

Preisler, S., Rebolj, M., Untermann, A., Ejegod, D. M., Lynge, E., Rygaard, C., & Bonde, J. (2013). Prevalence of human papillomavirus in 5,072 consecutive cervical SurePath samples evaluated with the Roche cobas HPV real-time PCR assay. *PLoS ONE*, *8*(3), e59765. doi:10.1371/journal.pone.0059765 PMID:23533648

Qureshi, M. N., Bolick, D., Ringer, P. J., Spagler, F. L., & Zimmerman, G. (2005). HPV testing in liquid cytology specimens: Comparison of analytic sensitivty and specificity for in situ hybridization and chemiluminescent nucleic acid testing. *Acta Cytologica*, *49*(2), 120–126. doi:10.1159/000326118 PMID:15839613

Rebolj, M., Bonde, J., Njor, S. H., & Lynge, E. (2011).Human papillomavirus testing in primary cervical screening and the cut-off level for hybrid capture 2 tests: Systematic review. *British Medical Journal*, *342*(d2757).

Rebolj, M., Preisler, S., Ejegod, D. M., Rygaard, C., Lynge, E., & Bonde, J. (2014).Disagreement between human papillomavirus assays: An unexpected challenge for the choice of an assay in primary cervical screening.[PubMed]. *PLoS ONE*, *9*(1), e86835. doi:10.1371/journal.pone.0086835

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Revzina, N. V., & Diclemente, R. J. (2005). Prevalence and incidence of human papillomavirus infection in women in the USA: A systematic review. *International Journal of STD & AIDS*, *16*(8), 528–537. doi:10.1258/0956462054679214 PMID:16105186

Rousseau, M. C., Pereira, J. S., Prado, J. C., Villa, L. L., Rohan, T. E., & Franco, E. L. (2001). Cervical coinfection with human papillomavirus (HPV) types as a predictor of acquisition and persistence of HPV infection. *The Journal of Infectious Diseases*, *184*(12), 1508–1517. doi:10.1086/324579 PMID:11740725

Sá, J. P. M. d. (2007). *Applied statistics: using SPSS, Statistica, MATLAB, and R*. New York: Springer.

Saiti, F., Naini, A. A., Shoorehdeli, M. A., & Teshnehlab, M. (2009). Thyroid Disease Diagnosis Based on Genetic Algorithms Using PNN and SVM. In *3rd International Conference on Bioinformatics and Biomedical Engineering*(pp. 1-4) doi:10.1109/ICBBE.2009.5163689

Scheffner, M., & Whitaker, N. J. (2003). Human papillomavirus-induced carcinogenesis and the ubiquitin-proteasome system. *Seminars in Cancer Biology*, *13*(1), 59–67. doi:10.1016/S1044-579X(02)00100-1 PMID:12507557

Siebers, A. G., Klinkhamer, P. J. J. M., Grefte, J. M. M., Massuger, L. F. A. G., Vedder, J. E. M., Beijers-Broos, A., & Arbyn, M. et al. (2009). Comparison of liquid-based cytology with conventional cytology for detection of cervical cancer precursors: A randomized controlled trial. *Journal of the American Medical Association*, *302*(16), 1757–1764. doi:10.1001/jama.2009.1569 PMID:19861667

Smith, J. H. (2002). Bethesda 2001. *Cytopathology*, *13*(1), 4–10. doi:10.1046/j.1365-2303.2002.00397.x PMID:11985563

Smits, H. L., van Gemen, B., Schukkink, R., van der Velden, J., Tjong, A. H. S. P., Jebbink, M. F., & ter Schegget, J. (1995). Application of the NASBA nucleic acid amplification method for the detection of human papillomavirus type 16 E6-E7 transcripts. *Journal of Virological Methods*, *54*(1), 75–81. doi:10.1016/0166-0934(95)00032-P PMID:7559859

Solomon, D., Davey, D., Kurman, R., Moriarty, A., O'Connor, D., Prey, M., & Young, N. et al. (2002). The 2001 Bethesda System: Terminology for reporting results of cervical cytology. *Journal of the American Medical Association*, *287*(16), 2114–2119. doi:10.1001/jama.287.16.2114 PMID:11966386

Sorbye, S. W., Arbyn, M., Fismen, S., Gutteberg, T. J., & Mortensen, E. S. (2011). HPV E6/E7 mRNA testing is more specific than cytology in post-colposcopy follow-up of women with negative cervical biopsy. *PLoS ONE*, *6*(10), e26022. doi:10.1371/journal.pone.0026022 PMID:21998748

Sorbye, S. W., Fismen, S., Gutteberg, T., & Mortensen, E. S. (2010). Triage of women with minor cervical lesions: Data suggesting a "test and treat" approach for HPV E6/E7 mRNA testing. *PLoS ONE*, *5*(9), e12724. PMID:20856930

Spathis, A., Kottaridi, C., Chranioti, A., Meristoudis, C., Chrelias, C., Panayiotides, I. G., & Karakitsos, P. et al. (2012). mRNA and DNA detection of human papillomaviruses in women of all ages attending two colposcopy clinics. *PLoS ONE*, *7*(11), e49205. doi:10.1371/journal.pone.0049205 PMID:23166611

Specht, D. F. (1990). Probabilistic neural networks and the polynomial Adaline as complementary techniques for classification. *IEEE Transactions on Neural Networks*, *1*(1), 111–121. doi:10.1109/72.80210 PMID:18282828

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Stevens, M. P., Garland, S. M., Rudland, E., Tan, J., Quinn, M. A., & Tabrizi, S. N. (2007). Comparison of the Digene Hybrid Capture 2 assay and Roche AMPLICOR and LINEAR ARRAY human papillomavirus (HPV) tests in detecting high-risk HPV genotypes in specimens from women with previous abnormal Pap smear results. *Journal of Clinical Microbiology*, *45*(7), 2130–2137. doi:10.1128/JCM.02438-06 PMID:17494721

Stoler, M. H., Wright, T. C. J., Sharma, A., Apple, R., Gutekunst, K., & Wright, T. L. (2011). High-risk human papillomavirus testing in women with ASC-US cytology: Results from the ATHENA HPV study. *American Journal of Clinical Pathology*, *135*(3), 468–475. doi:10.1309/AJCPZ5JY6FCVNMOT PMID:21350104

Szarewski, A., Ambroisine, L., Cadman, L., Austin, J., Ho, L., Terry, G., & Cuzick, J. et al. (2008). Comparison of predictors for high-grade cervical intraepithelial neoplasia in women with abnormal smears. *Cancer Epidemiology, Biomarkers & Prevention*, *17*(11), 3033–3042. doi:10.1158/1055-9965.EPI-08-0508 PMID:18957520

Szarewski, A., Mesher, D., Cadman, L., Austin, J., Ashdown-Barr, L., Ho, L., & Cuzick, J. et al. (2012). Comparison of seven tests for high-grade cervical intraepithelial neoplasia in women with abnormal smears: The Predictors 2 study. *Journal of Clinical Microbiology*, *50*(6), 1867–1873. doi:10.1128/JCM.00181-12 PMID:22422852

Terry, G., Ho, L., Londesborough, P., Cuzick, J., Mielzynska-Lohnas, I., & Lorincz, A. (2001). Detection of high-risk HPV types by the hybrid capture 2 test. *Journal of Medical Virology*, *65*(1), 155–162. doi:10.1002/jmv.2015 PMID:11505458

Trope, A., Sjoborg, K., Eskild, A., Cuschieri, K., Eriksen, T., Thoresen, S., & Lie, A. K. et al. (2009). Performance of human papillomavirus DNA and mRNA testing strategies for women with and without cervical neoplasia. *Journal of Clinical Microbiology*, *47*(8), 2458–2464. doi:10.1128/JCM.01863-08 PMID:19535524

Trottier, H., & Franco, E. L. (2006). The epidemiology of genital human papillomavirus infection. *Vaccine*, *24*(S1), 1–15.

Tsiodras, S., Hatzakis, A., Spathis, A., Margari, N., Meristoudis, C., Chranioti, A., … Karakitsos, P. (2011). Molecular epidemiology of HPV infection using a clinical array methodology in 2952 women in Greece. Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases.

Tsoumpou, I., Arbyn, M., Kyrgiou, M., Wentzensen, N., Koliopoulos, G., Martin-Hirsch, P., & Paraskevaidis, E. et al. (2009). (INK4a) immunostaining in cytological and histological specimens from the uterine cervix: A systematic review and meta-analysis. *Cancer Treatment Reviews*, *35*(3), 210–220.

Tyagi, S., Bratu, D. P., & Kramer, F. R. (1998). Multicolor molecular beacons for allele discrimination. *Nature Biotechnology*, *16*(1), 49–53. doi:10.1038/nbt0198-49 PMID:9447593

Tzivras, M., Megalopoulou, T. M., Pouliakis, A., Giahnaki, A. E., Karaitianos, I. G., Archondakis, S., & Karakitsos, P. et al. (2008). The potential of tree classifiers on the cytological examination of gastric lesions. *Archives of the Balkan Medical Union*, *43*(4), 239–244.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Ukpo, O. C., Flanagan, J. J., Ma, X.-J., Luo, Y., Thorstad, W. L., & Lewis, J. S. J. (2011). High-risk human papillomavirus E6/E7 mRNA detection by a novel in situ hybridization assay strongly correlates with p16 expression and patient outcomes in oropharyngeal squamous cell carcinoma. *The American Journal of Surgical Pathology*, *35*(9), 1343–1350. doi:10.1097/PAS.0b013e318220e59d PMID:21836494

Venturoli, S., Leo, E., Nocera, M., Barbieri, D., Cricca, M., Costa, S., & Zerbini, M. et al. (2012). Comparison of Abbott RealTime High Risk HPV and Hybrid Capture 2 for the detection of high-risk HPV DNA in a referral population setting. *Journal of Clinical Virology*, *53*(2), 121–124. doi:10.1016/j.jcv.2011.10.016 PMID:22115872

Walboomers, J. M., Jacobs, M. V., Manos, M. M., Bosch, F. X., Kummer, J. A., Shah, K. V., & Muñoz, N. et al. (1999). Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *The Journal of Pathology*, *189*(1), 12–19. doi:10.1002/(SICI)1096-9896(199909)189:1<12::AID-PATH431>3.0.CO;2-F PMID:10451482

Wallin, K. L., Wiklund, F., Angström, T., Bergman, F., Stendahl, U., Wadell, G., & Dillner, J. et al. (1999). Type-specific persistence of human papillomavirus DNA before the development of invasive cervical cancer. *The New England Journal of Medicine*, *341*(22), 1633–1638. doi:10.1056/NEJM199911253412201 PMID:10572150

Wasserman, P. D. (1993). Advanced methods in neural computing. New York: Van Nostrand Reinhold.

Wong, F. K., Ching, J. C., & Chow, J. K. (2010). Comparison of the DiagCor GenoFlow Human Papillomavirus Array Test and Roche Linear Array HPV Genotyping Test. *The Open Virology Journal, 4*, 169-174.

Wright, T. C. Jr, Stoler, M. H., Sharma, A., Zhang, G., Behrens, C., & Wright, T. L. (2011). Evaluation of HPV-16 and HPV-18 genotyping for the triage of women with high-risk HPV+ cytology-negative results. *American Journal of Clinical Pathology*, *136*(4), 578–586. doi:10.1309/AJCPTUS5EXAS6DKZ PMID:21917680

Wright, T. C. J. Jr, Massad, L. S., Dunton, C. J., Spitzer, M., Wilkinson, E. J., & Solomon, D. (2007). 2006 consensus guidelines for the management of women with abnormal cervical cancer screening tests. *American Journal of Obstetrics and Gynecology*, *197*(4), 346–355. doi:10.1016/j.ajog.2007.07.047 PMID:17904957

zur Hausen, H. (1976). Condylomata acuminata and human genital cancer. *Cancer Research*, *36*(2 pt 2), 794. PMID:175942

zur Hausen, H. (1991). Viruses in human cancers. *Science*, *254*(5035), 1167–1173. doi:10.1126/science.1659743 PMID:1659743

ADDITIONAL READING

Arbyn, M., Sasieni, P., Meijer, C. J., Clavel, C., Koliopoulos, G., & Dillner, J. (2006). Chapter 9: Clinical applications of HPV testing: a summary of meta-analyses. Vaccine, 24 Suppl 3, S3/78-89.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Bustin, S. A., & Nolan, T. (2004). Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. *Journal of Biomolecular Techniques*, *15*, 155–166. PMID:15331581

de Villiers, E. M., Fauquet, C., Broker, T. R., Bernard, H. U., & zur Hausen, H. (2004). Classification of papillomaviruses. *Virology*, *324*(1), 17–27. doi:10.1016/j.virol.2004.03.033 PMID:15183049

Feinberg, A. P., & Tycko, B. (2004). The history of cancer epigenetics. *Nature Reviews. Cancer*, *4*(2), 143–153. doi:10.1038/nrc1279 PMID:14732866

Gosden, R. G., & Feinberg, A. P. (2007). Genetics and epigenetics--nature's pen-and-pencil set. *The New England Journal of Medicine*, *356*(7), 731–733. doi:10.1056/NEJMe068284 PMID:17301306

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: The next generation. *Cell*, *144*(5), 646–674. doi:10.1016/j.cell.2011.02.013 PMID:21376230

Heid, C. A., Stevens, J., Livak, K. J., & Williams, P. M. (1996). Real time quantitative PCR. *Genome Research*, *6*(10), 986–994. doi:10.1101/gr.6.10.986 PMID:8908518

Hubbard, R. A. (2003). Human papillomavirus testing methods. *Archives of Pathology & Laboratory Medicine*, *127*, 940–945. PMID:12873165

Peto, J., Gilham, C., Fletcher, O., & Matthews, F. E. (2004). The cervical cancer epidemic that screening has prevented in the UK. *Lancet*, *364*(9430), 249–256. doi:10.1016/S0140-6736(04)16674-9 PMID:15262102

Roden, R., & Wu, T. C. (2006). How will HPV vaccines affect cervical cancer? *Nature Reviews. Cancer*, *6*(10), 753–763. doi:10.1038/nrc1973 PMID:16990853

Ronco, G., Dillner, J., Elfstrom, K. M., Tunesi, S., Snijders, P. J., Arbyn, M., & Meijer, C. J. et al. (2014). Efficacy of HPV-based screening for prevention of invasive cervical cancer: Follow-up of four European randomised controlled trials. *Lancet*, *383*(9916), 524–532. doi:10.1016/S0140-6736(13)62218-7 PMID:24192252

Trottier, H., & Franco, E. L. (2006). The epidemiology of genital human papillomavirus infection. *Vaccine*, *24*(Suppl 1), S1–S15. doi:10.1016/j.vaccine.2005.09.054 PMID:16406226

Woodman, C. B., Collins, S. I., & Young, L. S. (2007). The natural history of cervical HPV infection: unresolved issues. *Nat Rev Cancer, 7,* 11-22. doi:nrc2050 [pii]10.1038/nrc2050

zur Hausen, H. (2002). Papillomaviruses and cancer: From basic studies to clinical application. *Nature Reviews. Cancer*, *2*(5), 342–350. doi:10.1038/nrc798 PMID:12044010

This work was previously published in the Handbook of Research on Trends in the Diagnosis and Treatment of Chronic Conditions edited by Dimitrios I. Fotiadis, pages 99-132, copyright year 2016 by Medical Information Science Reference (an imprint of IGI Global).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 9
Proposed Threshold Algorithm for Accurate Segmentation for Skin Lesion

T. Y. Satheesha
Nagarjuna College of Engineering and Technology, India

D. Sathyanarayana
Ragiv Gandi Memorial College of Engineering and Technology, India

M. N. Giri Prasad
Jawaharlal Nehru Technological University, India

ABSTRACT

Automated diagnosis of skin cancer can be easily achieved only by effective segmentation of skin lesion. But this is a highly challenging task due to the presence of intensity variations in the images of skin lesions. The authors here, have presented a histogram analysis based fuzzy C mean threshold technique to overcome the drawbacks. This not only reduces the computational complexity but also unifies advantages of soft and hard threshold algorithms. Calculation of threshold values even the presence of abrupt intensity variations is simplified. Segmentation of skin lesions is easily achieved, in a more efficient way in the following algorithm. The experimental verification here is done on a large set of skin lesion images containing every possible artifacts which highly contributes to reversed segmentation outputs. This algorithm efficiency was measured based on a comparison with other prominent threshold methods. This approach has performed reasonably well and can be implemented in the expert skin cancer diagnostic systems.

1. INTRODUCTION

Image segmentation of an image is one of the most important and difficult task. The threshold is one of the most commonly used method for segmentation, it can also be used as the starting step for other segmentation methods. Classic segmentation algorithm class is a good time legible. However, in many

DOI: 10.4018/978-1-5225-0549-5.ch009

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

real-life situations, such as image segmentation of skin cancer, which is very hard to partition. Maintaining structural details of the diagnosis of skin cancer is important, because the lesion boundary provides information about the degree of asymmetry and steep are the details of a melanoma and benign lesions is important to distinguish between the functions.

Classic threshold scheme (Glasbey, 1993) distributes pixel clearly and in the region between the pixels may not be distinguished, even if their gray scale values are different in the original image. Therefore, the use of the structural details of the loss of a simple hard thresholding scheme, such as may result from such a simple hard threshold method does not seem to be good choice lesion segmentation. Replace conventional deterministic samples assigned to a class, the fuzzy division strategy (Aja-Fernandez, Vegas-Sanchez-Ferrero, & Fernandez, 2010) classes provided, wherein each sample point is assigned a soft description in the algorithm, with a given members of each class. Such partitioning reflects the identity of the sample, and thereby retains the structural details of the original data set. Identity is stored in the pixel blur partition space, since the distribution to members of the pixel depends on the difference between its grayscale value and the average gradation value of an area between it belongs.

Natural extension of fuzzy clustering segmentation considered separately as gray value features, resulting in the development of the threshold (Cheng & Chen, 1999). FCM (Fuzzy C-means) currently exist in the literature implement threshold (Fuzzy C-means) are mainly two main ways. The first is an extreme search for all possible threshold, while the second is based on a recognition by iterative evaluation of the optimal combination of standard features, and updates the threshold value. Two iterative and non-iterative implementation on the prospects (Jawahar, Biswas, & Ray, 2000) proposed a look at support efficient execution iteration, as long as it converges.

Pixel classification problem is formulated as a very suitable clustering problem. Traditional valley search options threshold required to provide the optimal threshold on Pattern Recognition sense. Classical deterministic optimal partition does not reflect the geometry of the intensity distribution, but this limitation is to take care of a fuzzy environment. Experimental details are discussed and validated the importance of fuzzy threshold scheme based on fuzzy clustering. Generally, however, the benefit is limited to an image analysis method to extract an object from one scene to a set of features characterizing objects. Even a vague description of the threshold would be sufficient for this purpose, the traditional feature extraction and object recognition methods may not be applicable to this specification as an example. Thus, although based fuzzy threshold (or above) describe the development of elegant image analysis techniques exist, hardening schemes require description can be used to make a conventional object recognition program. Some details of the hardening process, can be found in Jawahar et al. (2000) and Sezgin and Sankur (2004) in the literature. Close attention to the soft and hard threshold scheme different strengths, we are here, we propose a method based on the integration of these two types of threshold algorithm.

The main objective of this paper is to provide the major difference between this algorithm is popular in its implementation and iterative formula. Many iterative optimization algorithm instead, to an extensive search to identify the global maximum or minimum value associated with the target. Function. The proposed algorithm reduces the computational burden threshold selection basis using the histogram analysis done by means clustering fuzzy C. Compared to the same as reported in the literature Otsu method (Jung & Jinglu, 2008), the minimum error threshold (Morii, 1991), the minimum cross-entropy algorithm as hard algorithm will be discussed in the next section provides a better segmentation of lesions (El-Zaart et al., 2008) and keep Tsai (1985) and Kapoor entropy-based (Kapur, Sahoo, & Wong, 1985) as well as conventional soft threshold based on fuzzy C-means clustering.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

This paper is organized in the following format: Section 2 has the proposed algorithm and all details. Section 3 provides experimental results and discusses the performance and evaluation. The overall remarks are concluded at the end.

2. THE PROPOSED ALGORITHM

FCM clustering and histogram-based threshold has some distinct advantages. Histogram-based threshold value is based on the discovery hard threshold may be a good way mode image with well separated in the histogram. On the basis of the other fuzzy clustering and gray-scale images allowed to classify pixel intensity values of more than two classes. It's hard to unify and fuzzy threshold scheme approach is proposed in this paper.

Description existence elegant image analysis technology developed based on fuzzy clustering, plans must be made to strengthen the traditional description of a useful target recognition program. For the future with main idea of the algorithm from our observation, image analysis methods are often interested in limiting the objects extracted from the background, so that the object with a set of features characterization.

Most of the hardening techniques are limited to only 2-class fuzzy partitioning, e.g. (1) Finding a crisp threshold between two cluster centres, i.e. $U_1 < T_f < U_2$ and $\mu_o(T_f) = \mu_B(T_f)$. This provides mutually exclusive and exhaustive object and background regions. (2) Hardening based on α – cuts of fuzzy sets, such as $O = O_\alpha = \{(m,n); \mu_o(I\ mn) \geq \alpha\}$ and B= G-O. The parameter $\alpha \in (0,1]$ directly controls the size of the object region. As α increases, O approaches the core/skeleton of the object region. We found that not much attention is given for finding hardening schemes for general fuzzy segmentation, where the number of classes is more than two.

We proposed to use 3-class histogram analysis based fuzzy c mean thresholding which combines the benefits of fuzzy c mean clustering and histogram analysis methods.

At first, based on the intensity values of the pixels its broadly classified into three types. Now N objects are parted into C classes using the FCM clustering technique. The aim of FCM algorithm is to find an optimal fuzzy c-partition by evolving the fuzzy partition matrix U= [Ui,j] iteratively and computing the cluster centers.

In order to achieve this objective function the, algorithm tries to minimize the functionality:

$$q = \sum_{i=1}^{c} \sum_{j=1}^{N} \left(U_{i,j}\right)^{m} \left[z_j - v_i\right]^2$$

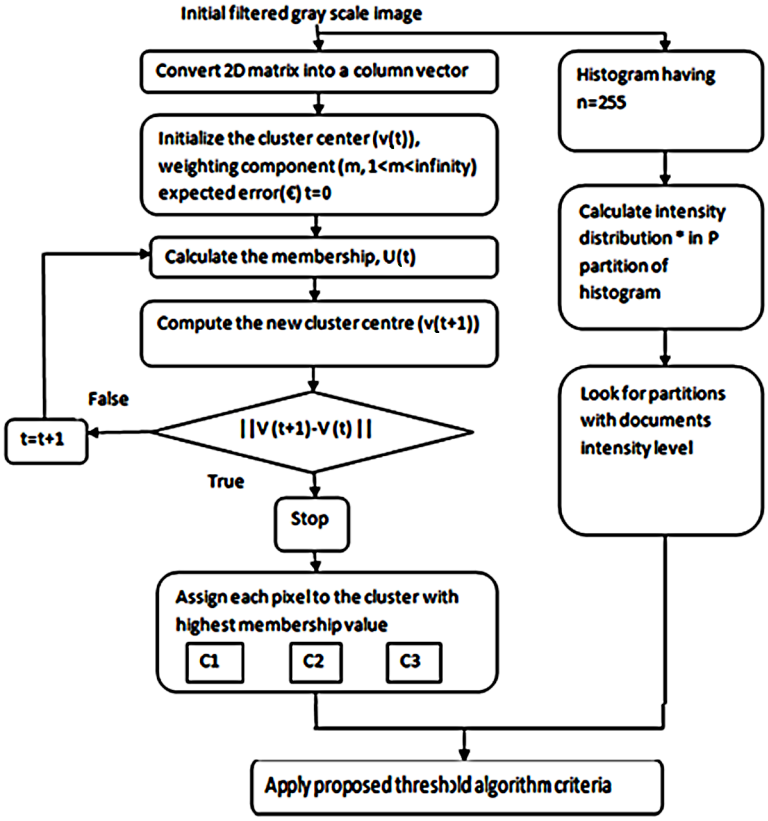
iteratively updating the cluster centers and the membership functions using the following Equations 1 and 2 (Masood & Al-Jumaily, 2013):

$$v_i = \left(\sum_{j=1}^{N} \left(u_{ij}\right)^m\right) \Big/ \left(Z_j \sum_{j=1}^{N} \left(u_{ij}\right)^m\right) \tag{1}$$

$$U_{ij} = 1 \Big/ \left(\sum_{k=1}^{C} \left(\left(v_i - u_j\right) \Big/ \left(v_i - u_k\right)\right)^{2/m-1}\right) \tag{2}$$

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 1. Proposed thresholding algorithm



After performing FCM clustering, finally each pixel is assigned to the cluster for which its membership value is maximum.

The proposed thresholding algorithm is presented in Figure 1. Histogram analysis of the filtered image is done to calculate for the intensity distribution in P partitions of histogram with P=C, where C is the number of clusters. For our application of skin images we found experimentally that C=3 provide best results. Then the clusters with dominant intensity value are used for finding the appropriate threshold value using the proposed threshold selection criterion.

We found that the actual meaning of multidimensional space of fuzzy clustering is very important when it comes to skin cancer diagnosis (Dongju & Jian, 2009). Increase the accuracy of the segmentation result in this case, the error is determined classification can be very expensive and complicated process required threshold. In the proposed method, the histogram of the image analysis is done to see an image of the intensity distribution of the mean, then the hard threshold chosen between classes and dominant intensity value. Therefore, this method may appear to solve the effective threshold automatically find different images, regardless of the strength of various difficulties. This method can also be used as the initialization step has spatial information (Dongju & Jian, 2009) a method for dividing the composite. In addition to skin cancer diagnosis, the preparation of the proposed high-level vision can provide additional useful information, as well as.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

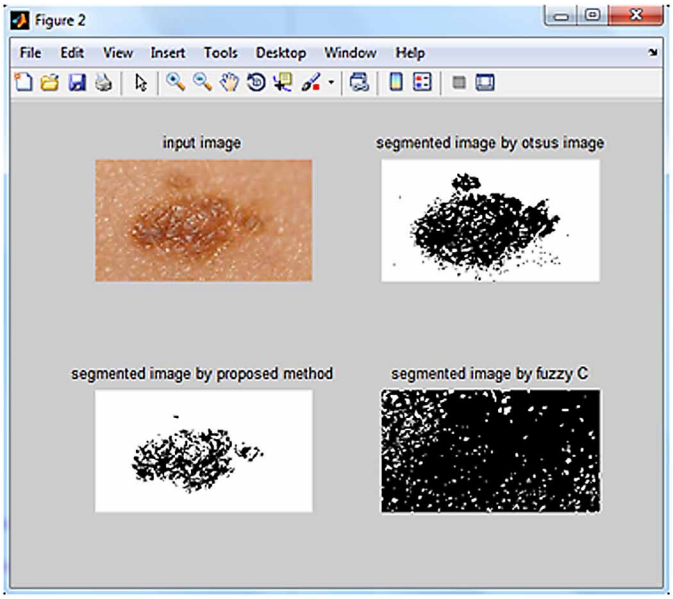
Table 1. Comparative analysis

Threshold Method	Evaluation Parameter		
	Coefficient of Similarity = a	**Border Error = b**	**Dice Coefficient = c**
Ostus	0.91	9.0	0.79
Fuzzy C mean method	0.90	9.7	0.81
Proposed method	0.96	3.1	0.98

Coefficient of similarity = a= 1 - |Vot-Vsrl/Vsr
Border error = b = (Area $_{sr}$ U Area$_{ot}$ - Area$_{sx}$ π Area $_{ot}$) / (Area $_{ot}$)
Dice coefficient = c = (2TP) / (((FP+TP) + (TP+FN)))
Where TP=true positive, FP=false positive, TN=true negative, FN=false positive.

The main difference between this algorithm and proposed algorithm is that it combines the advantages of fuzzy C-means clustering iterative formula and histogram analysis. Therefore, it reduces the computational burden in the traditional FCM threshold (Jawahar et al., 2000) require threshold iterative process. In addition to this it has overcome the search for the identification of appropriate threshold range, needs as required in the conventional histogram threshold (Masood & Al-Jumaily, 2013). Has been confirmed, the algorithm discussed in this article is effective, as a comparison, the literature reports algorithm (Sezgin & Sankur, 2004; Jung & Jinglu, 2008; El-Zaart, 2008; Sookpotharom, 2009; Ben Chaabane, 2008). Other thresholds provide superior segmentation results.

Figure 2. Threshold of images with non-melanoma lesion



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 3. Threshold of images with melanoma lesion

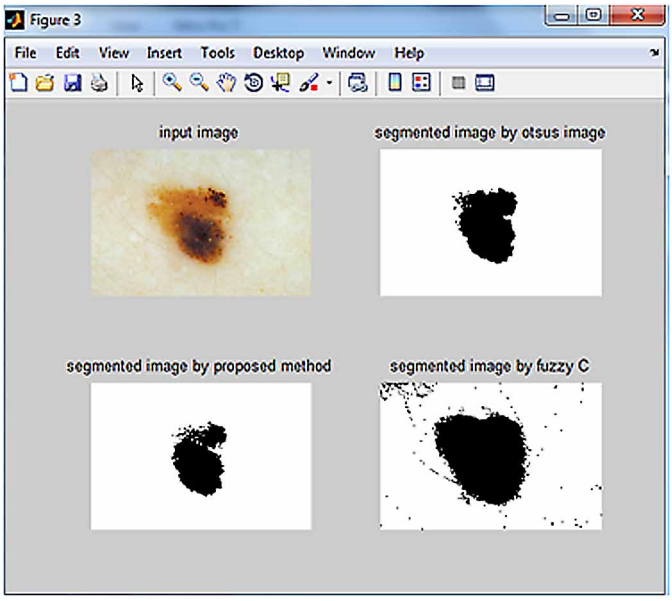
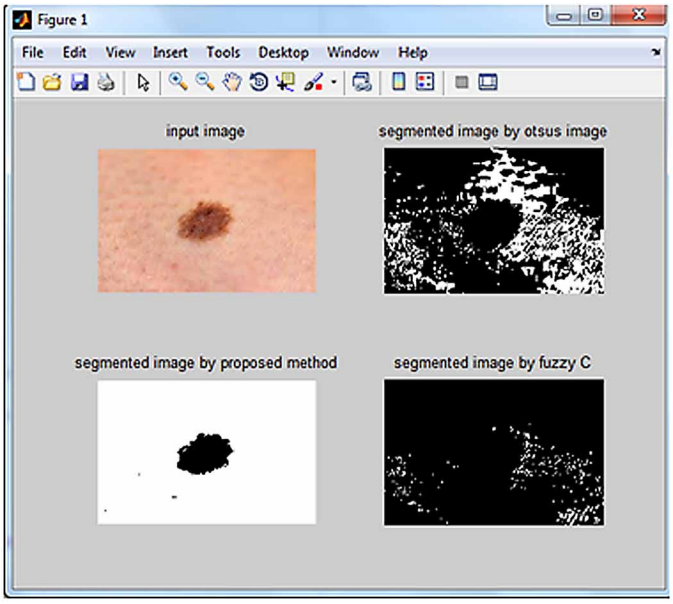


Figure 4. Threshold of images with benign lesion



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

3. EXPERIMENTAL EVALUATION

In order to establish the effectiveness of the algorithm and the clinical applicability, segmentation algorithm in the objective evaluation of a large set of clinical data is one important step. To evaluate the efficiency of the proposed method provided a method of comparing the performance of a known threshold value set some of the same skin lesions image segmentation used herein.

Comparison of the results of the proposed segmentation method with seven other known threshold method are shown in Table 1. They are used to analyze the collected from different sources, but most of the images obtained from the clinical view mirrors and 270 skin lesions submit an image database Sydney Melanoma Diagnostic Centre, Royal Prince Alfred Hospital. Segmentation results with the reference image (ground truth) and mean scores were compared segmentation are given in Table 1 for each method.

It is composed of the proposed method has been shown quite better performance compared to the results of other methods obvious. Results threshold treatment of certain skin lesions images are presented in Figure 2, Figure 3, and Figure 4. Other methods are required to apply, and the algorithm provides before even in the presence of a good hair and skin texture segmentation results of the threshold method of hair removal. Similarly, based on the maximum likelihood and entropy are not performing well for many colored lesions. On the other hand, Otsu's method, fuzzy C-means clustering and keep the threshold-based selection cannot deal with bad lighting image, if the lesion color is also very close to the skin color. Minimum cross entropy and minimum errors show relatively good performance, but still cannot reach the algorithm. It can be observed that the efficiency of the proposed algorithm provides most efficient thresholding results even in the presence of artifacts like hair, multiple colours inside lesion or ill-illumination etc.

4. CONCLUSION

Algorithms of thresholding method and histogram, combines the advantages of each of these methods, taking into account the overall simplicity of the algorithm. It takes into account the image histogram analysis, and automatically segments having different intensity distributions of the skin image. The algorithm can also provide the number of lesions with multiple images quite good results. A good initial segmentation is essential sophisticated methods to achieve global convergence faster. The result of this algorithm can be used as information for the use of space, as well as more sophisticated method of dividing a starting image.

REFERENCES

Aja-Fernandez, S., Vegas-Sanchez-Ferrero, G., & Martin Fernandez, M. A. (2010), "Soft thresholding for medical image segmentation," in *Proc. Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (pp. 4752-4755).

Ben Chaabane, S. (2008), Color image segmentation using automatic thresholding and the fuzzy C-means techniques. in *Proc. 14th IEEE Mediterranean Electro technical Conference, MELECON.(* pp. 857-861). doi:10.1109/MELCON.2008.4618543

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Cheng, H. D., & Chen, Y.-H. (1999). Fuzzy partition of two-dimensional histogram and its application to thresholding. *Pattern Recognition*, *32*(5), 825–843. doi:10.1016/S0031-3203(98)00080-6

Dongju, L., & Jian, Y. (2009), Otsu Method and K-means. in *Proc. Ninth International Conference on Hybrid Intelligent Systems*. (pp.344- 349).

El-Zaart, et. (2007), Thresholding of Medical Images Using Minimum Cross Entropy. in *International Federation for Medical and Biological Engineering (IFMBE) Conference*. Springer- Verlag Berlin Heidelberg, vol. 15, 296-299.

Glasbey, C. A. (1993). An Analysis of Histogram-Based Thresholding Algorithms *CVGIP*. *Graphical Models and Image Processing*, *55*(6), 532–537. doi:10.1006/cgip.1993.1040

Jawahar, C. V., Biswas, P. K., & Ray, A. K. (2000). Analysis of fuzzy thresholding schemes. *Pattern Recognition*, *33*(8), 1339–1349. doi:10.1016/S0031-3203(99)00122-3

Jun, Z., & Jinglu, H. (2008), Image Segmentation Based on 2D Otsu Method with Histogram Analysis. in *Proc. International Conference on Computer Science and Software Engineering*. (pp.105-108).

Kapur, J. N., Sahoo, P. K., & Wong, A. K. C. (1985). A new method for graylevel picture thresholding using the entropy of the histogram. *Computer Vision Graphics and Image Processing*, *29*(3), 273–285. doi:10.1016/0734-189X(85)90125-2

Masood, A., & al-Jumaily, A. A. (2013). Fuzzy C mean Thresholding based Level Set for Automated Segmentation of Skin Lesions. *Journal of Signal and Information Processing*, *4*(033B), 66–71. doi:10.4236/jsip.2013.43B012

Morii, F. (1991). A note on minimum error thresholding. *Pattern Recognition Letters*, *12*(6), 349–351. doi:10.1016/S0167-8655(05)80004-2

Sezgin, M., & Sankur, B. l. (2004). Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, *31*(1), 146–168.

Sookpotharom, S. (2009), "Border Detection of Skin Lesion Images Based on Fuzzy C-Means Thresholding". in *Proc. 3rd International Conference on Genetic and Evolutionary Computing*. (pp. 777-780). doi:10.1109/WGEC.2009.96

Tsai, W.-H. (1985). Moment-preserving thresolding: A new approach. *Computer Vision Graphics and Image Processing*, *29*(3), 377–393. doi:10.1016/0734-189X(85)90133-1

This work was previously published in the International Journal of Biomedical and Clinical Engineering (IJBCE), 4(2); edited by N. Sriraam, pages 40-47, copyright year 2015 by IGI Publishing (an imprint of IGI Global).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Section 2
Technologies and Applications

This section presents an extensive coverage of various tools and technologies available in the field of oncology and discusses a variety of applications and opportunities available that can be considered by practitioners in developing viable and effective oncology programs and processes. With 11 chapters, this section offers a broad treatment of some of the many tools and technologies within the oncology field. These chapters enlighten readers about fundamental research on the many tools facilitating the burgeoning field of oncology and review topics from case studies to best practices and ongoing research. Further chapters discuss oncology in a variety of settings. Contributions included in this section provide excellent coverage of today's IT community and how research into oncology is impacting the social fabric of our present-day global village. It is through these rigorously researched chapters that the reader is provided with countless examples of the up-and-coming tools and technologies emerging from the field of oncology.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 10
Nanotechnology, Metal Nanoparticles, and Biomedical Applications of Nanotechnology

M. Amin Bhat
Sathyabama University, India

Anima Nanda
Sathyabama University, India

B. K. Nayak
KMC PGS Autonomus Pudicherry, India

Imtiyaz H. Lone
IGNOU Maidan Garhi, India

ABSTRACT

Nanotechnology has emerged as an important field of modern scientific research due to its diverse range of applications in the area of electronics, material sciences, biomedical engineering, and medicines at nano levels such as healthcare, cosmetics, food and feed, environmental health, optics, biomedical sciences, chemical industries, drug-gene delivery, energy science, optoelectronics, catalysis, reprography, single electron transistors, light emitters, nonlinear optical devices, and photoelectrochemical applications and other applications. Due to these immense applications of nanotechnology in biomedical science, it has became possible to design the pharmaceuticals in such a way that they could directly treat diseased cells like cancer and make microscopic repairs in hard-to-operate-on areas of the body. The nanomachines have been designed to clean up toxins or oil spills, recycle all garbage, eliminate landfills, etc. The chapter summarizes the present and future applications of nanotechnology for human welfare but needs further study in catalysis, optical devices, sensor technology, cancer treatment, and drug delivery systems.

INTRODUCTION

Nowadays technical and engineering activities have a significantly greater impact on the natural laws and rules once considered as changeless. Nonetheless we have been approaching the same pattern that we followed in the past – we get inspired by the nature. Nanoscience is currently a fast growing field and Nanotechnology is at the cutting edge of this rapidly evolving area (Mandal et al., 2006). Nanomaterials and nanotechnologies include molecules and atoms, biosensors and multisensory systems measure the

DOI: 10.4018/978-1-5225-0549-5.ch010

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

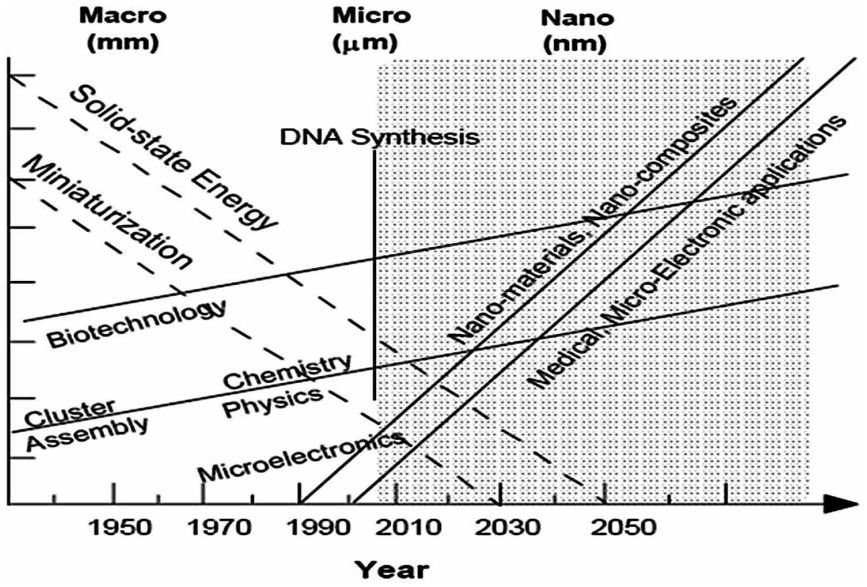
EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

required biological parameters and processes with a relatively high accuracy, the electronic systems and computers create artificial intelligence that is integrated into diagnostic, measuring or control systems etc likely to change the way almost everything from vaccines to computers to automobile tires to objects not yet imagined is designed and made. Nanotechnology collectively describes technology and science involving nano-scale particles (nanoparticles) that increases the scope of investigating and regulating the interplay at cell level between synthetic materials and biological systems (Du et al., 2007), which can be applied as an efficient tool to explore the finest process in biological processes (Sondi & Salopak, Sondi, 2004) and in biomedical sciences (Hutten et al., 2004).

The word "nano" means dwarf in Greek language and is used to indicate one billionth of a meter or 10-9 and a world of things is built up from the tiny scale of nanometers. The term nanotechnology was coined by Norio Taniguchi a, researcher at the University of Tokyo, Japan. Nano science and technology is a broad and interdisciplinary area of research and development activity that has been growing explosively worldwide in the past few years and making an impact in all spheres of human-life (Vaidyanathanh et al., 2009). Besides this, It has the potential for revolutionizing the ways in which materials and products are created and the range and nature of functionalities that can be accessed. Nanotechnology plays an indispensable role in drug delivery, diagnostics, imaging, sensing, gene delivery, artificial implants and tissue engineering, has significant commercial impact, which will be assuredly increased in the future. (Marones et al., 2004).

The nanotechnology has also shown major developments in is the production and application of nanoparticles in biology. Novel techniques and methods are being are constantly being studied to produce nanoparticles. The enormous interest has developed green synthesis of NPs is due to their unusual optical (Krotiknowska et al., 2003), chemical (Kumar et al., 2003), Photochemical (Chandrasekharan et al., 2000), electronic (Peto et al., 2002) and magnetic (Watson et al., 1999), properties. In the chapter we would like the emphases the importance of metallic nanoparticles synthesised through various methods and the other applications of nanotechnology with respect to biomedical applications.

Figure 1. Evolution of science and technology and the future



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

HISTORICAL BACKGROUND

The history of nanoparticle research very old and the use of these particles date back to the 9th century in Mesopotamia when artisans used these to generate a glittering effect on the surface of pots. This lustre or glitter over pottery from the middle Ages and Renaissance is due to a metallic film that was applied to the transparent surface of a glazing. The lustre can still be visible if the film has resisted atmospheric oxidation and other weathering. The lustre is within the film itself which contained silver and copper nanoparticles dispersed homogeneously in the glassy matrix of the ceramic glaze. Artisans created the nanoparticles by adding copper and silver salts and oxides together with vinegar, ochre and clay, on the surface of previously-glazed pottery. Then the pots were placed into a kiln and heated to about 600 °C in a reducing atmosphere. With the heat the glaze would soften, causing the copper and silver ions to migrate into the outer layers of the glaze. Michael Faraday 1857 firstly proved scientifically that optical property of nanometals. Nanoparticles are either newly created via nanotechnology or are present naturally over the earth's crust or in the environment caused by weathering of Au deposits. The metal formed by evaporation is coupled with minerals and been deposited rapidly from saline ground water. Nanotechnology has been defined as a technology that mainly consist of the process of separation, consolidation and deformation of materials by one atom or molecule (Taniguchi, 1974). The following are the mile stones in the evolution of Nanotechnology.

3.5 Mrd. years: First cells with nano-machines.
400 B.C: Demokrit: Reasoning about atoms and matter.
1905: Albert Einstein: Calculated molecular diameter.
1931: Max Knoll, Ernst Ruska: Electron microscope.
1959: Richard Feynman: There is plenty of room at the bottom.
1968: Alfred Y.Cho, JohnArthur (Bell Labs): MBE (atomic layer growth).
1974: Norio Taniguchi: Nanotechnology for fabrication methods below1μm.
1981: Gerd Binnig and Heinrich Rohrer: Noble prize for inventing STM.
1985: Robert F.Carl, Harald W.Kroto: Richard Smalley: Bucky balls.
1986: K. Eric Drexler: writing with a STM tool.
1991: Sumio Ligima: Carbon Nanotubes.
1993: Warren Robinett, R.Stanley Williams: Combination of SEM and VR (virtual reality system).
1998: Cees Dekkar et al.: Carbon nano-tube transistor.
1999: James M. Tour and Mark A. Read: Single molecule switch.
2000: Eigler et al.: Construction of quantum mirrors.
2001: Florian Bambers: Soldering of nanotube with e-beam.
2004: Intel launches the Pentium IV 'PRESCOFT' processor based on 90nm technology.

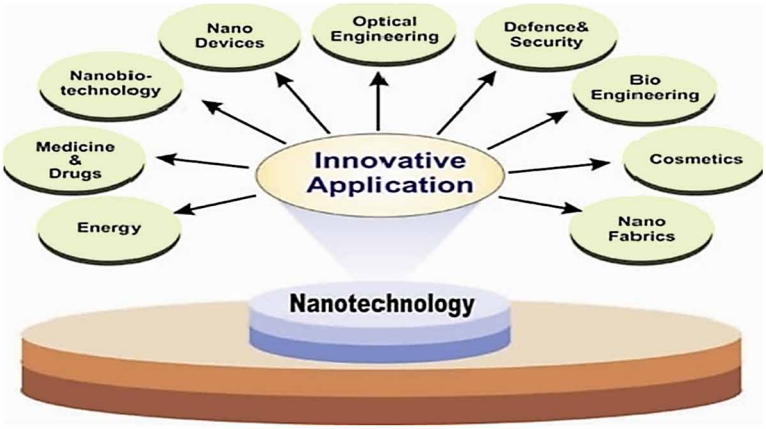
The metals like Silver, gold, Titanium, Magnesium, Selenium, Iron, Tin etc were known only as a metal till recently and it is only when the nano-era came into existence that people started to believe that could even be produced at the nanoscale. The Nanoparticles were being used by various organisms for bioremediation of various metals before the existence of nanotechnology it remained unnoticed. The number of organisms has been used to purify environment by removing various reactive metal salts from the environment. Both the Gram positive and Gram negative metals have been used for the biosorption of metals such as silver, cadmium and copper. Various organisms have been tested for their efficiency

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

to adsorb these metal ions and these metal ions remained as a colloidal aggregate mostly on the cell surface, occasionally on the cytoplasm. The size of the nanoparticles played no significant role during those times. These capable organisms were in turn termed as the "competent organisms" which could bind large amounts of metal ions (Beveridge and Murray 1976; Mullen et al. 1989; Doyle et al. 1980). Cell walls of Gram positive bacteria such as *Bacillus subtilis* were found to bind with large quantities of metals than the Gram negative bacteria such as *Escherichia coli* (Beveridge and Fyfe 1985). The synthesis methods in the early 1980s described the reduction of metal ion as a two step procedure; in first step very small particles were synthesized, which were then enlarged to several nanometres. The difference remained in the use of the reducing agent for the synthesis where in the former step a stronger reducing agent was used, and in latter step a weaker reducing agent was used (Sintubin et al. 2009). Chemical and physical methods were used for the synthesis of nanoparticles, a controlled process mediated by the addition of specific reducing agents at raised temperatures and various pH values. Metal nanoparticles were synthesized through an array of methods such as spark discharging, electrochemical reduction, solution irradiation and cryochemical synthesis but the biosynthesis of nanoparticles has emerging and highlighted the intersection of nanotechnology and biotechnology attention due to a growing need to develop environmentally benign technologies in material synthesis (Kalish waralal et al. 2008).

The role of microorganisms for the production of nanoparticles in the fate of metals in the environment was not thoroughly examined; however, it was conceived that they represent an important component of metal dynamics. The first evidence of biosynthesis metal nanoparticles (silver) was established in 1984 using the microorganism *Pseudomonas stutzeri* AG259, a bacterial strain that was originally isolated from silver mine (Haefeli et al. 1984; Zhang et al. 2005; Nair and Pradeep 2002). The deposition of silver on bacterial cell membrane under the transmission electron microscope was confirmed which opened new avenues towards the preparation of nanostructural materials that incorporate metal-based crystalline particles with defined structural, compositional and morphological properties. Biological methods are gaining importance because of the use of normal conditions for the synthesis that enable control over the size of the nanoparticles. The biological synthesis of nanoparticles germinated from the experiments on biosorption of metals bacteria were not identified as nanoparticles but were named as aggregates (Mullen et al. 1989). The microorganisms, such as bacteria, yeast and fungi play an important role in remediation

Figure 2. Nanotechnology over view



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

of toxic metals through reduction of the metal ions; this was considered interesting as nanofactory, using the dissimilatory properties of fungi and hence these biological systems have been extensively used for the rapid and eco-friendly biosynthesis of metal nanoparticles (Bhattacharya and Gupta, 2005). Unicellular as well multi-cellular organisms have shown immense potential for the synthesis of nanoparticles, as they can synthesize inorganic materials either intracellularly or extracellularly. More recent and detailed investigations into the use of microbes in the synthesis of different metal nanoparticles include bacteria for gold (Beveridge and Murray, 1980; Southam and Beveridge, 1996; Fortin and Beveridge, 2000), silver (Joerger et al., 2000; Klaus et al,.1998), ZnS (Labrenz et al., 2000), magnetite (Lovley et al., 1987; Nair and Pradeep, 2002) CdS (Smith et al., 1998; Philipse and Maas,2002), iron sulfide (Watson et al., 1999, 2000), yeast for PbS (Kowshik et al., 2002a), CdS (Kowshik et al., 2002b), and silver (Kowshik et al., 2003), and algae for gold (Robinson et al., 1997). *Morganella* sp. (Parikh et al. 2008).

SYNTHESIS OF NANOPARTICLES

The synthesis and study nanoparticles, has become a major interdisciplinary area of research over the past 10 years. The size, morphology as well as the properties of nanoparticles basically depends on the methods of preparation. All the processes can be broadly divided into three processes, physical methods, chemical methods and biological methods.

Physicochemical Methods

Nanoparticles are mainly prepared by physical and chemical methods. Metallic nanoparticles, including gold, silver, iron, zinc and metal oxide nanoparticles, have shown great promise in biomedical application, due to their large surface area to volume ratio are being produced by both methods. The silver nanoparticles can also be prepared by both methods called ‘top-down’ and ‘bottomup’ methods. The top-down method mainly deals with the mechanical grinding of bulk metals and subsequent stabilization of the subsequent nanosized metal particles by the addition of colloidal protecting agents (Gaffet E et al., 1996, Amulyavichus A et al., 1998). While as bottom-up methods deal with reduction of metals by electrochemical methods, and sonodecomposition. There are also numerous physical and chemical approaches for synthesis of silver nanoparticles reported

- Spray pyrolysis.
- Gas condensation.
- Freeze drying.
- Ultrasonic methods.
- Chemical precipitation.
- Sol-gel technique.
- Polymeric precursor.
- Micro emulsion.
- Hydrothermal synthesis.
- Combustion synthesis.
- Dc Arc Plasma Process.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

- Rf Plasma Process.
- Plasma Rapid Solidification Technology.
- Reactive Electrode Submerged Arc.

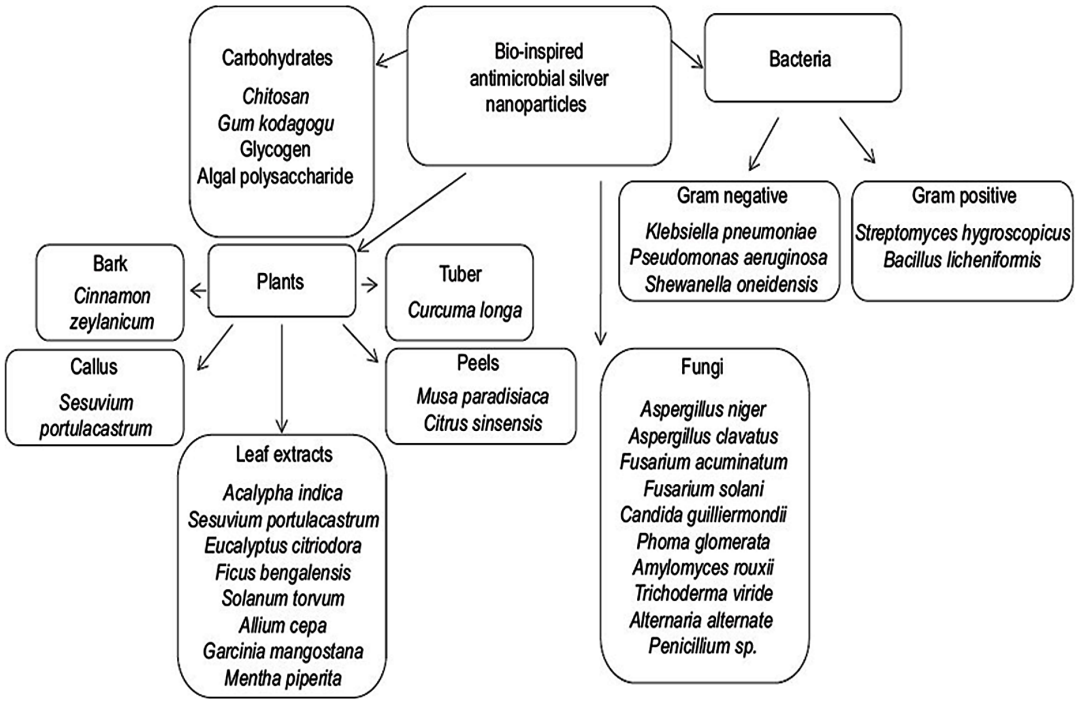
Biological Methods

The metallic nanoparticles prepared through the chemical and physical methods are costly and also include the use of toxic chemicals, thus there is a need for an environmentally and economically feasible way to synthesize these nanoparticles. The metallic nanoparticles can be synthesised by biological methods which are easy to synthesise, cost effective, ecofriendly and non toxic (Mukherjee et al., 2001) is possible with help of bacteria, fungi, and plant extracts. Biological synthesis of silver nanoparticles is a bottom-up method that includes reduction/oxidation reactions. The microbial enzymes or photochemical with antioxidant or reducing properties act on the corresponding compounds and give the anticipated nanoparticles (Kalishwaralal, et al., 2008, Parashar, UK et al., 2009).

BIOSYNTHESIS OF METALLIC NANOPARTICLES USING BACTERIA

The bacteria are most abundant microorganisms which are prokaryotic in nature and greatly diversified in size, shape, and lives in all kinds of habitats, including extreme environments that exhibit, for example, extremely high or low temperatures, acidity, alkalinity, and salt or sulfur concentrations are known as

Figure 3. Bioinspired nanoparticles



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

extremophile. Some bacterial species of bacteria have exhibited the ability to resort to specific defense mechanisms to quell stresses like toxicity of foreign metal ions or metals; even at high ion concentrations. The *Pseudomonas stutzeri* is living in soil rich in silver (Haefeli et al., 1984) and *Pseudomonas aeruginosa* have developed resistance to that metal (Bridges et al., 1979). Some bacteria are known to synthesize inorganic materials like the magnetotactic bacteria, which synthesize intracellular magnetite NPs (Lovley et al., 1987; Spring and Schleifer, 1995). Consequently, a wide number of bacterial species have been used in nanobiotechnology to research alternative methods for the production of metallic nanoparticles (Table 1).

Table 1. List of bacterial strains utilised for the biosynthesis 0f metallic nanoparticles

Bacteria	Nanoparticles	Shape	Average Size (nm)	Type of Microscopic Analysis	Reference
Pseudomonas stutzeri	Ag	Triangles and hexagons	200	TEM, HRTEM	Klaus et al. 1999)
Actinobacter spp.	Fe3 O4	Quasi-spherical, cubic	10–40, 50–150	TEM	Bharde et al. (2005)
Rhodobacter sphaeroides	ZnS	Spherical	4, 8, 30 and 105	TEM	Bai et al. (2006)
Shewanella oneidensis	Au	Spherical	12 ± 5	TEM	Suresh et al. (2011)
Shewanella algae	Pt	NR	5	TEM	Konishi et al. (2007)
Thermoanaerobacter ethanolicus	Fe3 O4	Octahedral	>35	TEM, SEM	Roh et al. (2007)
Bacillus sphaericus	Pd	NR, clusters of sc	1, sc 0.6–0.8	SEM	Pollmann et al. (2006)
Plectonema boryanum	Au	Cubic, irregular, octahedral	10–25, <10, 10–25	TEM, SEM	Lengke et al. (2006a)
Shewanella sp.	Se	Spherical	181, 164, 68, 50	TEM, SEM	Lee et al. (2007)
Pseudomonas aeruginosa	Au	Quasi-spherical	40, 25, 15	TEM	Husseiny et al. (2007)
Aeromonas sp.	Ag	Clusters	20	TEM, SEM	Zhang et al. (2007)
Escherichia coli	Au	Spheres	5–50, <10	TEM	Deplanche and Macaskie -2008
Recombinant	CdZn, CdSe, CdTe, SeZn	NR	10.35, 4.99	TEM, HRTEM, CM	Park et al. (2010)
Desulfovibrio desulfuricans	Au	Spheres	5–50, <10	TEM	Deplanche and Macaskie (2008)
	Pd	NR	NR	TEM	Deplanche et al. (2008)
	Au	NR	20–50	TEM	Deplanche et al. (2008)
Thermoanaerobacter sp.	Mx Fe1−x OOH (M = Zn, Co, Mn and Ni)	NR	25, 51, 53, 75	TEM	Moon et al (2007)
Rhodopseudomonas capsulata	Au	Nanowires	50–60	TEM	He et al. (2008)

cointinued on following page

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 1. Continued

Bacteria	Nanoparticles	Shape	Average Size (nm)	Type of Microscopic Analysis	Reference
Bacillus licheniformis	Ag	NR	50	SEM	Kalimuthu et al. (2008)
	Au	Cubes	10–100	SEM	Kalishwaralal et al. -2009
	Ag	NR	50	TEM	Vaidyanathan et al. -2010
	Au	NR	10–50	TEM	Kalishwaralal et al. (2010b)
Proteus mirabilis	Ag	Spherical	10–20	TEM	Samadi et al. (2009)
Bacillus cereus	Ag	Spherical	4–5	TEM, HRTEM	Ganesh Babu and Gunasekaran (2009)
Morganella sp.	Ag	Spherical	20 ± 5	TEM, HRTEM	Parikh et al. (2008)
Staphylococcus aureus	Ag	Agglomerates	160–180	AFM	Nanda and Saravanna (2009)
Bacillus subtilis	Au	Spherical	7.6 ± 1.8, 7.3 ± 2.3	TEM	Reddy et al. (2010)
Bacillus thuringiensis	Ag	Variable	15	TEM	Jain et al. (2010)
Mycobacterium smegmatis	Cu	Doughnut shape	5, 10–20	AFM	Wörner et al. (2007)
Streptomyces	Ag	Spherical	20–30	TEM, SEM, AFM	Sadhasivam et al. (2010)
Hygroscopicus	Au	Various	Various	HRTEM, FESEM, AFM	Sadhasivam et al. (2012)
Streptomyces albidoflavus	Ag	Spherical	10–40	TEM	Prakasham et al. (2012)

Abbreviations: NR, Not Reported; CS, Cellular Surface; PS, Periplasmic Space; CM, Cytoplasmic Membrane; CFE, Cell Free Extract; ICF, Inactive Cell Filtrate, SCM, Spore crystal mixture; PCS, Phytochelatin Synthease, MT, Metallothionein FESEM, Field Emission Scanning Electron Microscopy: SEM, Scanning Electron microscopy; TEM, Transmission electron microscopy; HRTEM, High Resolution Transmission Electron Microscopy; EELS, electron energy-loss spectroscopy; STEM, Scanning Transmission Microscopy; SPM, Kelvin-Probe/Surface Potential Microscopy; AFM, Atomic Force Microscopy; FM, Fluorescence microscopy; CM, Confocal microscopy. EFTEM-ESI, Energy Filtering Transmission Electron Microscopy-Element Specific Imaging

BIOSYNTHESIS OF METALLIC NANOPARTICLES USING FUNGI

The fungi are eukaryotic organisms, mostly decomposers, having more than 70,000 species living in a wide variety of habits. As per the recent estimated data, there are as many as 5.1 million fungal species exist (Blackwell, 2011). These organisms possess the ability to digest food extracellularly, secreting special enzymes to hydrolyze complex compounds into simpler molecules, which are then absorbed and used as an energy source. Fungi are excellent candidates to explore their use in nanobiotechnology, because some species grow at fast rates and are very easy to culture and maintain in the laboratory (Castro-Longoria et al., 2011, 2012). The use of fungal species as manufactories for the production of nanoparticles has attained a great interest of nanotechnologists recently; as firstly fungal species has been utilised for bio-synthesis of silver nanoparticles from the filamentous fungus *Verticillium sp*. (Mukherjee et al., 2001). The number of fungal species have been exploited their after for the production of nanoparticles. Mostly

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

the biomass is commonly exposed to metallic ion solutions, extracellular synthesis of nanoparticles (Ahmad et al.2002). The number of fungal species have been reported for CdS nanoparticles as well as the formation of Ag, Au, PbS, ZnS, and MoS_2 etc. The presence of proteins in the aqueous solution suggested a possible sulfate reducing enzyme-based process for production of nanoparticles. Depending on the metallic ion solution and incubation conditions, diverse morphologies can be obtained, although quasi-spherical nanoparticles are the most commonly produced nanoparticles. From different species as extracellular and intracellular formation has been reported presenting different size ranges as well as varying shapes. The exact mechanism involved in synthesis of metallic nanoparticles using fungi is not clear. But it is believed that the mechanism involve in synthesis of metallic nanoparticles by fungi is trapping of Ag^+ ions at the surface of the fungal cells and the subsequent reduction of the metal ions by the enzymes present in the fungal system. Fungi can produce larger amounts of nanoparticles as compared to bacteria as they can secrete larger amounts of proteins which directly translate to higher yield of nanoparticles. *F. oxysporum*, it is believed that the NADPH-dependent nitrate reductase process is responsible for nanoparticle formation (Mohanpuria et al., 2008, Mukherjee et al., 2001, and Ahmad et al., 2003) *Penicillium glabrum*(MTCC-1985),(Nanda & Shahnaz, 2014), *Penicillium sp* (Nayak et al., 2013) and *Aspergillus Sp* (Nayak et al., 2014). Different fungi species have been exploited for the biosynthesis of metallic nanoparticles have been reported in Table 2.

Table 2. List of fungal strains utilized for the biosynthesis of metallic nanoparticles

	Nanoparticles	Shape	Size (nm)	Type of Microscopic Analysis	Reference
Verticillium	Ag	Spherical	25	SEM, TEM	Mukherjee et al.(2001)
Fusarium oxysporum	Ag	Highly variable	5–15	TEM	Ahmad et al. (2002)
	Au	Spherical and triangular	8–40	TEM	Mukherjee et al. (2002)
	Au	Spherical, triangular	46, 21c	TEM	Shankar et al. (2004a)
	Ag	Spherical	20–50	SEM	Durán et al. (2005)
Fusarium oxysporum	Magnetite	Irregular, quasi-spherical	20–50	TEM	Bharde et al. (2006)
Aspergillus fumigatus	Ag	Mostly spherical, some triangular	5–25	TEM	Bhainsa and D'Souza (2006)
	Ag	Mostly spherical, rectangular	15–45	TEM	Alani et al. (2012)
Lycopersici Phaenerochaete	Ag	Varying, mainly pyramidal	50–200	SEM, TEM	Vigneshwaran et al. (2006)
Chrysosporium Pleurotus sajor caju	Ag	Spherical	30.5	TEM	Vigneshwaran et al. (2007a)
Aspergillus flavus	Ag	Spherical	8.92	SEM, TEM	Vigneshwaran et al. (2007b)
	TiO_2	Spherical	62–74	SEM, TEM, AFM	Rajakumar et al. (2012)
Aspergillus niger	Au	Nonoplates	50–500	FESEM	Xie et al. (2007c)
Trichoderma asperellum	Ag	NR	13–18	TEM, HRTEM	Mukherjee et al. (2008)
Volvariella volvaceae	Ag	Spherical	15	TEM, HRTEM	Philip (2009)
Penicillium fellutanum	Ag	Mostly spherical	5–25	TEM	Kathiresan et al. (2009)
Cladosporium cladosporioides	Ag	Mostly spherical	10–100	TEM	Balaji et al. (2009)

cointinued on following page

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 2. Continued

	Nanoparticles	Shape	Size (nm)	Type of Microscopic Analysis	Reference
Phoma glomerata	Ag	Spherical	60–80	SEM	Birla et al. (2009)
Aspergillus versicolor	Hg	NR	NR	SEM, AFM, FM	Das et al. (2009a)
Trichoderma viride	Ag	Spherical, rodlike	5–40	TEM	Fayaz et al. (2009a)
Fusarium spp.	Zn	Irregular, some spherical	100–200	SEM, TEM	Velmurugan et al. (2010)
Aspergillus clavitus	Ag	NR	550–650	AFM	Saravanan and Nanda (2010)
Aspergillus flavus NJP08	Ag	Spherical	17	TEM, HRTEM	Jain et al. (2011)
Aspergillus flavus	Ag	Spherical	55-70	AFM, FESEM	Nayak et al (2014)
Aspergillus oryzae var.	Au	Spheres	10–60	TEM, FESEM	Binupriya et al. (2010b)
Streptomyces sp.	Ag	Mostly spherical	15–25	TEM	Alani et al. (2012)
Aspergillus terreus	Ag	Spherical	1–20	TEM	Li et al. (2012)
Alternaria alternata	Au	Spherical, triangular	12 ± 5	AFM, TEM	Sarkar et al. (2012)
Yeast					
Candida glabrata	CdS	NR	A° NR	EFTEM-ESI	Krumov et al. (2007)

Abbreviations: NR, Not Reported; CS, Cellular Surface; PS, Periplasmic Space; CM, Cytoplasmic Membrane; CFE, Cell Free Extract; ICF, Inactive Cell Filtrate, SCM, Spore crystal mixture; PCS, Phytochelatin Synthease, MT, Metallothionein FESEM, Field Emission Scanning Electron Microscopy: SEM, Scanning Electron microscopy; TEM, Transmission electron microscopy; HRTEM, High Resolution Transmission Electron Microscopy; EELS, electron energy-loss spectroscopy; STEM, Scanning Transmission Microscopy; SPM, Kelvin-Probe/Surface Potential Microscopy; AFM, Atomic Force Microscopy; FM, Fluorescence microscopy; CM, Confocal microscopy. EFTEM-ESI, Energy Filtering Transmission Electron Microscopy-Element Specific Imaging

BIOSYNTHESIS OF METALLIC NANOPARTICLES USING PLANTS

The plants and plant extracts have been explored under different environmental conditions to produce metallic nanoparticles. Synthesize metallic nanoparticles using microbes is very slow process when in comparison with plant extracts. The major advantage of using plant extracts for metallic nanoparticles synthesis is that they are easily available, safe, and nontoxic in most cases, have a broad variety of metabolites that can aid in the reduction of silver ions, and are quicker than microbes in the synthesis. Plant extract is the most commonly product used for the production of metallic nanoparticles. It has been proved through experiments that in some cases live plants show the ability to absorb metals from the outside and accumulate them in their tissues in the form of nanostructures (Gardea-Torresdey et al., 2002, 2003; Armendariz et al., 2004; Bali et al., 2006; Harris and Bali, 2008). The plant alfalfa (*Medicago sativa*) extract have been initially tested and distinctive metallic nanostructures were synthesized under diverse environmental conditions under varying pH, show differences in nanoparicle shapes and sizes are achieved (Ascencio et al., 2003, 2005; Canizal et al., 2006; Schabes-Retchkiman et al., 2006; Herrera-Becerra et al., 2007). The possible application of metallic nanoparticles synthesized using plants or plant extracts has not been entirely explored. The major phytochemical mechanism involved is the reduction of the ions involved are terpenoids, flavones, ketones, aldehydes, amides, and carboxylic acids. Flavones, organic acids, and quinones are water-soluble phytochemicals that are responsible for the immediate reduction of the ions. Studies have revealed that xerophytes contain emodin, an anthraquinone that undergoes tautomerization, leading to the formation of the silver nanoparticles. In the case of mesophytes, it was

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

found that they contain three types of benzoquinones: cyperoquinone, dietchequinone, and remirin. It was suggested that the phytochemicals are involved directly in the reduction of the ions and formation of metallic nanoparticles. Metallic nanoparticles synthesized using plant is shown in Table 3.

Types of Nanoparticles

Extensive libraries of nanoparticles, composed of an assortment of different sizes, shapes, and materials, and with various chemical and surface properties, have already been constructed. The field of nanotechnology is under constant and rapid growth and new additions continue to supplement these libraries. The classes of nanoparticles listed below are all very general and multi-functional, however, some of their basic properties and current known uses in biotechnology, and particularly nanomedicine, are described here.

Table 3. List of plants utilized for the biosynthesis 0f metallic nanoparticles

Plant	Nanoparticles	Shape	Size(nm)	Type of analysis	Source or Reference
Medicago sativa	Au	NR, images presented	2–20, 20–40	TEM, HRTEM	Gardea-Torresdey et al.(2002) (2002)b
	Ag	Icosahedral	2–20, 2–4	TEM, HRTEM	Gardea-Torresdey et al. (2003)
Pelargonium graveolens	Au	Decahedral, icosahedral	20–40	TEM	Shankar et al. (2003a)
	Ag	Spherical	16–40	TEM	Shankar et al. (2003b)
Azadirachta indica	Au	Spherical, triangular, hexagonal	NR	TEM	Shankar et al. (2004b)
Avena sativa	Au	Tetrahedral, decahedral, multitwinned, irregular, rod-shaped	5–20, 25–85	TEM	Armendariz et al. (2004)
Brassica juncea	Ag, Cu, Co, Zn, Ni	Spherical (Ag)	<100	TEM	Bali et al. (2006)a
	Ag	NR	NR	TEM	Harris and Bali (2008)
Aloe vera	Au	Triangular	50–350	TEM, AFM	Chandran et al. (2006)
	Ag	Spherical	15	TEM	Chandran et al. (2006)
Cinnamon camphora	Au	Spherical, triangular	10–40, 15–25	TEM, AFM	Huang et al. (2007)
	Ag	Spherical	55–80	TEM	Huang et al. (2007)
Coriandrum sativum	Au	Spherical, triangular, truncated triangular, decahedral	7–58	TEM	Badri Narayanan and Sakthivel (2008)
Cinnamon zeylanicun	Ag	Bud-like decahedral	31–40	TEM	Sathishkumar et al. (2009)
Magnolia kobus and Diopyros kaki	Au	Triangular, pentagonal, hexagonal, spherical	5–300	TEM, SEM AFM	Song et al. (2009)
Hibiscus rosa sinensis	Au	Spherical, triangular, hexagonal, dodecahedral	14	TEM, HRTEM	Philip (2010)
	Ag	Spherical	13	TEM, HRTEM	Philip (2010)
Citrus sinensis (peel extract)	Ag	Spherical	35 ± 2 (25 ∘ C), 10 ± 1 (60 ∘ C)	SEM, TEM	Kaviya et al., 2011

cointinued on following page

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 3. Continued

Plant	Nanoparticles	Shape	Size(nm)	Type of analysis	Source or Reference
Medicago sativa (seed exudates)	Ag	Hexagonal, triangular	5–100	SEM, TEM, AFM	Lukman et al. (2011)
Rhizophora mucronata	Ag	NR	60–95	AFM	Gnanadesigan et al. (2011)
Mimosa pudica	Ag	Spherical	25–60	SEM, TEM	Marimuthu et al. (2011)
Citrullus colocynthis (L-) Schrad	Ag	Spherical	31	AFM	Satyavani et al. (2011)
Pandanus odorifer (Forssk.)	Ag	Quasi-spherical	24–55	TEM	Panda et al. (2011)
Punica granatum	Ag	Spherical	5	TEM	Ahmad et al. (2012)
(peel extract)	Au	Spherical	10	TEM	Ahmad et al. (2012)
Citrus paradisi	Au	Mainly quasi-spherical	30	TEM, FESEM, CM	Nolasco-Arismendi et al. (2012)
Coccinia grandis	Ag	Spherical	20–30	SEM, TEM	Arunachalam et al. (2012)
Cassia fistula	Au	Various	55.2–98.4	SEM	Daisy and Saipriya (2012)
Musa paradisiaca	Ag	Rod	60–150	SEM	Jayaseelan et al. (2012)
Annona squamosa	Ag	Irregular spherical	35 ± 5	TEM	Kumar et al. (2012b)
(peel extract)	TiO_2	Spherical	23 ± 2	SEM, TEM	Roopan et al. (2012)
Ficus racemosa	Ag	Clindrical, uniform rod-shaped	250.6	SEM	Velayutham et al., 2013

Abbreviations: NR, Not Reported; CS, Cellular Surface; PS, Periplasmic Space; CM, Cytoplasmic Membrane; CFE, Cell Free Extract; ICF, Inactive Cell Filtrate, SCM, Spore crystal mixture; PCS, Phytochelatin Synthease, MT, Metallothionein FESEM, Field Emission Scanning Electron Microscopy: SEM, Scanning Electron microscopy; TEM, Transmission electron microscopy; HRTEM, High Resolution Transmission Electron Microscopy; EELS, electron energy-loss spectroscopy; STEM, Scanning Transmission Microscopy; SPM, Kelvin-Probe/Surface Potential Microscopy; AFM, Atomic Force Microscopy; FM, Fluorescence microscopy; CM, Confocal microscopy. EFTEM-ESI, Energy Filtering Transmission Electron Microscopy-Element Specific Imaging

1. **Fullerenes:** Buckyballs and Carbon tubes. Both members of the fullerene structural class, buckyballs and carbon tubes are carbon based, lattice-like, potentially porous molecules. Buckyballs are spherical in shape while carbon tubes are cylindrical. The diameter of a carbon tube can be several nanometer but the length can be much greater, up to several micrometer, depending on its intended use. Carbon nanotubes have many applications in materials science due to their strength and unique electrical properties. However, they have also found use in the field of biomedicine as carriers for vaccines, drugs and other molecules. A single wall carbon tube is a one-atom-thick sheet of graphite, resembling chicken wire, rolled seamlessly into a tube. There are also multi-walled and other types of tubes depending on the shape, diameter, density (hollow versus solid) and other properties.
2. **Liposomes:** Liposomes are lipid-based nanoparticles used extensively in the pharmaceutical and cosmetic industries because of their capacity for breaking down inside cells, once their delivery function has been met. Liposomes were the first engineered nanoparticles used for drug delivery but problems such as their propensity to fuse together in aqueous environments and release their payload, have lead to replacement, or stabilization using newer alternative nanoparticles.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

3. **Nanoshells:** Also referred to as core-shells, nanoshells are spherical cores of a particular compound surrounded by a shell or outer coating of another, which is a few nanometres thick. One application in biomedicine is to create nanoshells that absorb at biologically useful wavelengths, depending on the shell thickness. One common formula for the construction of nanoshells is to use silica for the core and another sticky compound to adhere gold particles to the outside surface, creating the shell. Nanoshells such as these have been used to kill cancer cells in mice. Once injected into a tumor, radiation is applied and the nanoshells heat up enough to kill the tumour cells.
4. **Dendrimers:** Dendrimers are highly branched structures gaining wide use in nanomedicine because of the multiple molecular "hooks" on their surfaces that can be used to attach cell-identification tags, fluorescent dyes, enzymes and other molecules. The first dendritic molecules were produced around 1980, but interest in them has blossomed more recently as biotechnological uses are discovered. Nanomedical applications for dendrimers are many and include nanoscale catalysts and reaction vessels, micelle mimics, imaging agents and chemical sensors, and agents for delivering drugs or genes into cells. There are two basic structural types. One is the globular structure with a central core from which branches radiate. The second type has no central core and consists simply of a series of highly branched polymers.
5. **Quantum Dots:** Also known as nanocrystals, quantum dots are nanosized semiconductors that, depending on their size, can emit light in all colours of the rainbow. These nanostructures confine conduction band electrons, valence band holes, or excitons in all three spatial directions. Examples of quantum dots are semiconductor nanocrystals and core-shell nanocrystals, where there is an interface between different semiconductor materials. They have been applied in biotechnology for cell labelling and imaging, particularly in cancer imaging studies.
6. **Super Paramagnetic Nanoparticles:** Super paramagnetic molecules are those that are attracted to a magnetic field but do not retain residual magnetism after the field is removed. Nanoparticles of iron oxide with diameters in the 5-100 nm range have been used for selective magnetic bioseparations. Typical techniques involve coating the particles with antibodies to cell-specific antigens, for separation from the surrounding matrix. Used in membrane transport studies, super paramagnetic iron oxide nanoparticles (SPION) are applied for drug delivery and gene transfection. Targeted delivery of drugs, bioactive molecules or DNA vectors is dependent on the application of an external magnetic force that accelerates and directs their progress towards the target tissue. They are also useful as MRI contrast agents.
7. **Nanorods:** Typically 1-100 nm in length, nanorods are most often made from semiconducting materials and used in nanomedicine as imaging and contrast agents. Nanorods can be made by generating small cylinders of silicon, gold or inorganic phosphate, among other materials. Current concerns over the safety of nanoparticles have lead to the development of many new facets of research. As a result, our collection of knowledge about nanoparticle interactions within cells is still rapidly growing.

BIOMEDICAL APPLICATIONS

In recent years, nanoparticles, coated and composite nanoparticles has been used in various fields of science, engineering and industrial applications by virtue of their specific, novel and useful properties. The unique properties of nanomaterials encourage belief that they can be applied in a wide range

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

of fields, from medical applications to environmental sciences. Studies conducted by nanotechnology experts mapping the risks and opportunities of nanotechnology have revealed enormous prospects for progress in both life sciences and information technology. Medical applications are expected to increase our quality of life through early diagnosis and treatment of diseases, and prosthetics, among others.

Nanoscaffolds

Nanofiber scaffolds can be used to regenerate central nervous system cells and possible other organs. Experiments performed on a hamster with severed optic tract demonstrated the regeneration of axonal tissue initiated by a peptide nanofibers scaffold (Ellis-Behnke et al., 2006).

Antimicrobial Nanopowders and Coatings

Certain nanopowders, possess antimicrobial properties (Bosi et al., 2003, Koper et al., 2002). When these powders contact cells of *E. coli*, or other bacteria species and viruses, over 90% are killed within a few minutes. Due to their antimicrobial effect, nanoparticle of silver and titanium dioxide (<100nm) are assessed as coatings for surgical masks (Li et al. 2006).

Bioseparation and Biosensors

Nanotube membranes can act as channels for highly selective transport of molecules and ions between solutions that are present on both side of the membrane (Jirage et al., 1997). For example, membranes containing Nanotubes with inside diameters of molecular dimensions (less than 1 nm) separate small molecules on the basis of molecular size, while Nanotubes with larger inside diameters (20– 60 nm) can be used to separate proteins (Martin et al., 2003). Biosensor is a device for the detection of an analyte that combines a biological component with a physicochemical detector component.

NANOROBOTICS AND DRUG DELIVERY

Nanorobotics is the technology of creating machines or robots at or close to the scale of a nanometer (10-9 meters). More specifically, nanorobotics refers to the still largely theoretical nanotechnology engineering discipline of designing and building nanorobots. Nanorobots (nanobots or nanoids) are typically devices ranging in size from 0.1-10 micrometers and constructed of nanoscale or molecular components. As no artificial non-biological nanorobots have so far been created, they remain a hypothetical concept at this time. Another definition sometimes used is a robot which allows precision interactions with nanoscale objects, or can manipulate with nanoscale resolution. Nanomedicine is the medical application of nanotechnology and related research. It covers areas such as nanoparticle drug delivery and possible future applications of molecular nanotechnology (MNT) and nanovaccinology.The ability of nanoparticles to target and penetrate specific organs and cells contributes to their toxicity; however, this ability may be exploited in nanomedicine. Nanospheres composed of biodegradable polymers can incorporate drugs, allowing the timed release of the drug as the polymer degrades (Uhrich et al., 1999). When particles are set to degrade in an acid microenvironment, such as tumor cells or around inflammation sites, this allows site-specific or targeted drug delivery.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Gene Transfection

Surface-functionalized nanoparticles can be used to permeate cell membranes at a much higher level than nanoparticles without a functionalized surface (Maïté et al. 2000). This property can be used to deliver genetic material into living cells, a process called transfection. For example, silica nanospheres labelled on their outer surfaces with cationic ammonium groups can bind DNA (a polyanion) through electrostatic interactions (Kneuer et al. 2000). Then nanoparticles deliver the DNA into cells.

Medical Imaging

A variety of techniques currently called "noninvasive" have been used for more than a quarter of a century in medical imaging, for example super paramagnetic magnetite particles coated with dextran are used as image-enhancement agents in magnetic resonance imaging (Harisinghani et al., 2003). Intracellular imaging is also possible through attachment of quantum dots to selected molecules, which allows intracellular processes to be observed directly.

Nasal Vaccination

Nanospheres carriers for vaccines are in development. Antigen-coated polystyrene nanospheres, used as vaccine carriers targeting human dendritic cells, have been researched for nasal vaccination (Matsusaki et al., 2005). Nanospheres had a direct effect on human dendritic cells, inducing transcription of genes important for, e.g., phagocytosis as well as an immune response.

Nucleic Acid Sequence and Protein Detection

Targeting and identifying various diseases could be made possible by detecting nucleic acid sequences unique to specific bacteria and viruses, or to specific diseases or abnormal concentration of certain proteins that signal the presence of various cancers and diseases Nanomaterials-based assays are currently evaluated as well as more sensitive proteins detections methods. Nucleic acid sequences are currently detected with polymerase chain reaction (PCR) coupled with molecular fluorophore assays. Despite high sensitivity, PCR has significant drawbacks, such as: complexity, sensitivity to contamination, cost, and lack of portability. Current protein detection methods, such as enzyme-linked immune absorbent assay (ELISA), allow the detection of proteins concentrations at which the disease is often advanced. More sensitive methods based on nanomaterials would revolutionize physical treatment of many cancer types and diseases (Rosi et al., 2005).

Smart Nanophase Extractors

Differentially functionalized nanotubes are used as smart nanophase extractors, with molecularrecognition capabilities, to remove specific molecules from solutions (Martin & Kohli, 2003).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Treatment for Local Anaesthetic Toxicity

Local anaesthetic can be sometimes very toxic, ranging from local neurotoxicity tocardiovascular collapse and coma. In addition to conventional therapies, drug-scavenging nanoparticles have shown to increase survival rate from no animals in the control group to all animals in the treated group (Renehan et al., 2005; Weinberg et al., 2003).

Protein Microarray

It is a piece of glass on which different molecules of protein have been affixed at separate locations in an ordered manner thus forming a microscopic array.

Nanomaterials in Water Treatment

Nanosorbents are nanoscale particles of inorganic or organic materials that are capable of absorbing other substances. Most environmental applications of nanosorbents are in the field of waste water treatment and drinking water production, with other applications focused on air pollutants or ground water contamination.

1. **Remediation:** Remediation is the science of removal or reduction of pollutants from the environment using chemical or biological means. Nanomaterials have been used for remediating contaminated groundwater and subsurface source areas of contamination at hazardous waste sites. Early treatment remedies for groundwater contamination were mainly pump-and-treat operations Due to the considerably high cost and long operating periods, the use of in-situ treatment technologies is considerably increasing.
2. **Bioremediation:** Bioremediation of harmful organic contaminants like carbazole and dibenzothiophene has been successfully reported by different groups using variety of microorganisms.

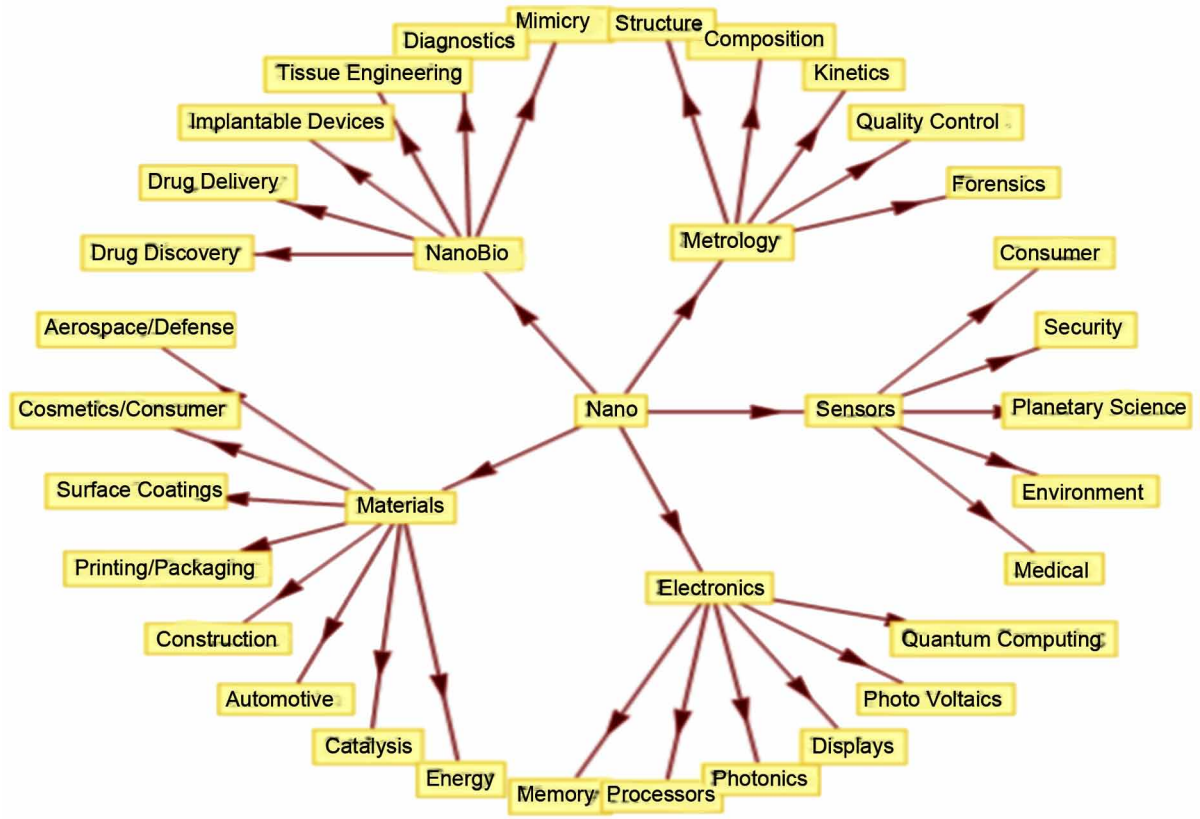
There are also numbers of applications of nanotechnology in various fields of science Figure 4, which have revolutionized the research world via various ways needs further studies.

CONCLUSION

Nanotechnology is an emerging field that is potentially changing the way we treat diseases through its diverse range of applications like Nanomedicines to control microbial pathogens, Nanorobotics and drug delivery which is having great achievement to treat targeted diseases. The tissue engineering and scaffolds nanotechnology, has made it possible to develop tissues to restore maintain and enhance tissue functions(langer and Vacanti 1993) and construction of three dimensional scaffolds from biomaterials which provide mechanical support and guide into new tissues and organs. The bacterial spore based Nano-vaccines has been tested against various diseases like tetanus with heterologous antigens and other diseases which have shown good applications eradicate the particular diseases. Quantam dots have been used to detect the malfunctioned cells which became possible through nanotechnology. Carbon nanotubes and biosensors have revolutionised the biomedical industry like regular self testing for glucose level,

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 4. Applications of nanotechnology in various fields



and could be used to monitor the pulse temperature diagnosing diseases etc. In the near future, the application of clean, non-toxic, and eco-friendly nanostructured material will be possible in industry and/or biomedicine. The other applications of nanotechnology like medical imaging, nucleic acid sequence and protein detection, water treatment has achieved a great success from last decade. As Feynman has predicted that there has been plenty of room at the bottom to modify and enhance existing technologies by controlling material properties at the nanoscale. This branch promises to further grow and explore many more possible applications, e.g. in catalysis, optical devices, sensor technology, cancer treatment, or drug delivery systems Therefore, with sufficient time and research the nanotechnology based biomedical applications are becoming a reality.

REFERENCES

Ahmad, A., Mukherjee, P., Mandal, D., Senapati, S., Khan, M. I., Kumar, R., & Sastry, M. (2002). Enzyme mediated extracellular synthesis of CdS nanoparticles by the fungus, Fusarium oxysporum. *Journal of the American Chemical Society*, *124*(41), 12108–12109. doi:10.1021/ja027296o PMID:12371846

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Ahmad, A., Mukherjee, P., Senapati, S., Mandal, D., Khan, M. I., Kumar, R., & Sastry, M. (2003). Extracellular biosynthesis of silver nanoparticles using the fungus Fusarium oxysporum. *Colloids and Surfaces. B, Biointerfaces*, *28*(4), 313–318. doi:10.1016/S0927-7765(02)00174-1

Ahmad, N., Sharma, S., & Rai, R. (2012). Rapid green synthesis of silver and gold nanoparticles using peels of *Punica granatum. Advanced Materials Letters*, *3*(5), 376–380.

Alani, F., Moo-Young, M., & Anderson, W. (2012). Biosynthesis of silver nanoparticles by a new strain of *Streptomyces sp*. Compared with Aspergillus fumigatus. *World Journal of Microbiology & Biotechnology*, *28*(3), 1081–1086. doi:10.1007/s11274-011-0906-0 PMID:22805829

Amulyavichus, A., Daugvila, A., Davidonis, R., & Sipavichus, C. (1998). Study of chemical composition of nanostructural materials prepared by laser cutting of metals. *Fizika Met Met.*, *85*, 111–117.

Armendariz, V., Herrera, I., Peralta-Videa, J. R., Jose-Yacaman, M., Troiani, H., Santiago, P., & Gardea-Torresdey, J. L. (2004). Size controlled gold nanoparticle formation by Avena sativa biomass: Use of plants in nanobiotechnology. *Journal of Nanoparticle Research*, *6*(4), 377–382. doi:10.1007/s11051-004-0741-4

Arunachalam, R., Dhanasingh, S., Kalimuthu, B., Uthirappan, M., Rose, C., & Mandal, A. B. (2012). Phytosynthesis of silver nanoparticles using *Coccinia grandis* leaf extract and its application in the photocatalytic degradation. *Colloids and surfaces. B, Biointerfaces*, *94*, 226–230. doi:10.1016/j.colsurfb.2012.01.040 PMID:22348986

Bai, H. J., Zhang, Z. M., & Gong, J. (2006). Biological synthesis of semiconductor zinc sul- fide nanoparticles by immobilized Rhodobacter sphaeroides. *Biotechnology Letters*, *28*(14), 1135–1139. doi:10.1007/s10529-006-9063-1 PMID:16794769

Bali, R., Razak, N., Lumb, A., & Harris, A. T. (2006). The synthesis of metal nanoparticles inside live plants. *Nanosci. Nanotechnol. International Conference, IEEE Xplore* DOI doi:10.1109/ICONN.2006.340592

Beveridge, T. J., & Fyfe, W. S. (1985). Metal fixation by bacterial cell walls. *Canadian Journal of Earth Sciences*, *22*(12), 1893–1898. doi:10.1139/e85-204

Beveridge, T. J., & Murray, R. G. E. (1976). Uptake and retention of metals by cell walls of Bacillus subtilis. *Journal of Bacteriology*, *127*, 1502–1518. PMID:821933

Beveridge, T. J., & Murray, R. G. E. (1980). Site of metal deposition in the cell wall of *Bacillus subtilis. Journal of Bacteriology*, *141*, 876–887. PMID:6767692

Bhainsa, K. C., & D'Souza, S. F. (2006). Extracellular biosynthesis of silver nanoparticles using the fungus Aspergillus fumigatus. *Colloids and surfaces. B, Biointerfaces*, *47*(2), 160–164. doi:10.1016/j.colsurfb.2005.11.026 PMID:16420977

Bharde, A., Rautaray, D., Bansal, V., Ahmad, A., Sarkar, I., & Yusuf, S. M. et al. (2006). Extracellular biosynthesis of magnetite using fungi. *Small*, *2*(1), 135–141. doi:10.1002/smll.200500180 PMID:17193569

Bharde, A., Wani, A., Shouche, Y., Joy, P. A., Prasad, B. L. V., & Sastry, M. (2005). Bacterial synthesis of nanocrystalline magnetite. *Journal of the American Chemical Society*, *127*(26), 9326–9327. doi:10.1021/ja0508469 PMID:15984833

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Bhattacharya, D., & Gupta, R. K. (2005). Nanotechnology and potential of microorganisms. *Critical Reviews in Biotechnology*, *25*(4), 199–204. doi:10.1080/07388550500361994 PMID:16419617

Binupriya, A. R., Sathishkumar, M., Vijayaraghavan, K., & Yun, S.-I. (2010). Bioreduction of trivalent aurum to nano-crystalline gold particles by active and inactive cells and cell-free extract of Aspergillus oryzae var. viridis. *Journal of Hazardous Materials*, *177*(1-3), 539–545. doi:10.1016/j.jhazmat.2009.12.066 PMID:20056324

Birla, S. S., Tiwari, V. V., Gade, A. K., Ingle, A. P., Yadav, A. P., & Rai, M. K. (2009). Fabrication of silver nanoparticles by *Phoma glomerata* and its combined effects against Escherichia coli, Pseudomonas aeruginosa and Staphylococcus aureus. *Letters in Applied Microbiology*, *48*(2), 173–179. doi:10.1111/j.1472-765X.2008.02510.x PMID:19141039

Blackwell, M. (2011). The fungi: 1, 2, 3... 5.1 million species? *American Journal of Botany*, *98*(3), 426–438. doi:10.3732/ajb.1000298 PMID:21613136

Bosi, S., da Ros, T., Spalluto, G., & Prato, M. (2003). Fullerene derivatives: An attractive tool for biological applications. *European Journal of Medicinal Chemistry*, *38*(11-12), 913–923. doi:10.1016/j.ejmech.2003.09.005 PMID:14642323

Bridges, K., Kidson, A., Lowbury, E. J. L., & Wilkins, M. D. (1979). Gentamicin and silver- resistant Pseudomonas in a burns unit. *British Medical Journal*, *1*(6161), 446–449. doi:10.1136/bmj.1.6161.446 PMID:106914

Chandran, S. P., Chaudhary, M., Pasricha, R., Ahmad, A., & Sastry, M. (2006). Synthesis of gold nanotriangles and silver nanoparticles using Aloe vera plant extract. *Biotechnology Progress*, *22*(2), 577–583. doi:10.1021/bp0501423 PMID:16599579

Chandrasekharan, N., & Kamat, P. V. (2000). Improving the photo-electrochemical performance of nanostructured TiO2 films by adsorption of gold nanoparticles. *The Journal of Physical Chemistry B*, *104*(46), 10851–10857. doi:10.1021/jp0010029

Daisy, P., & Saipriya, K. (2012). Biochemical analysis of Cassia fistula aqueous extract and phytochemically synthesized gold nanoparticles as hypoglycemic treatment for diabetes mellitus. *International Journal of Nanomedicine*, *7*, 1189–1202. doi:10.2147/IJN.S26650 PMID:22419867

Das, S. K., Das, A. R., & Guha, A. K. (2009). Adsorption behavior of mercury on function- alized Aspergillus versicolor mycelia: Atomic force microscopic study. *Langmuir*, *25*(1), 360–366. doi:10.1021/la802749t PMID:19115871

Deplanche, K., Caldelari, I., Mikheenko, I. P., Sargent, F., & Macaskie, L. E. (2010). Involvement of hydrogenases in the formation of highly catalytic Pd(0) nanoparticles by bioreduction of Pd(II) using Escherichia coli mutant strains. *Microbiology*, *156*(9), 2630–2640. doi:10.1099/mic.0.036681-0 PMID:20542928

Deplanche, K., & Macaskie, L. E. (2008). Biorecovery of gold byEscherichia coli andDesulfovibrio desulfuricans. *Biotechnology and Bioengineering*, *99*(5), 1055–1064. doi:10.1002/bit.21688 PMID:17969152

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Deplanche, K., Woods, R. D., Mikheenko, I. P., Sockett, R. E., & Macaskie, L. E. (2008). b. Manufacture of stable palladium and gold nanoparticles on native and genetically engineered flagella scaffolds. *Biotechnology and Bioengineering*, *101*(5), 873–880. doi:10.1002/bit.21966 PMID:18819156

Du, L., Jiang, H., Liu, X., & Wang, E. (2007). Biosynthesis of gold nanoparticles assisted by Escherichia coli DH5α and its application on direct electrochemistry of hemoglobin. *Electrochemistry Communications*, *9*(5), 1165–1170. doi:10.1016/j.elecom.2007.01.007

Durán, N., Marcato, P. D., Alves, O. L., De Souza, G. I. H., & Esposito, E. (2005). Mechanistic aspects of biosynthesis of silver nanoparticles by several Fusarium oxysporum strains. *Journal of Nanobiotechnology*, 8. PMID:16014167

Ellis-Behnke, R. G., Liang, Y. X., You, S. W., Tay, D. K., Zhang, S., So, K. F., & Schneider, G. E. (2006). Nano neuro knitting: Peptide nanofibers scaffold for brain repair and axon regeneration with functional return of vision. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(13), 5054–5059. doi:10.1073/pnas.0600559103 PMID:16549776

Fayaz, M., Tiwary, C. S., Kalaichelvan, P. T., & Venkatesan, R. (2010). Blue orange light emission from biogenic synthesized silver nanoparticles using *Trichoderma viride. Colloids and surfaces. B, Biointerfaces*, *75*(1), 175–178. doi:10.1016/j.colsurfb.2009.08.028 PMID:19783414

Fortin, D., & Beveridge, T. J. (2000). In E. Baeuerien (Ed.), *Biomineralization From Biology to Biotechnology and Medical Applications* (p. 7). Weinheim: Wiley-VCH.

Gaffet, E., Tachikart, M., El Kedim, O., & Rahouadj, R. (1996). Nanostructural materials formation by mechanical alloying: Morphologic analysis based on transmission and scanning electron microscopic observations. *Materials Characterization*, *36*(4-5), 185–190. doi:10.1016/S1044-5803(96)00047-2

Ganesh Babu, M. M., & Gunasekaran, P. (2009). Production and structural characterization of crystalline silver nanoparticles from Bacillus cereus isolate. *Colloids and surfaces. B, Biointerfaces*, *74*(1), 191–195. doi:10.1016/j.colsurfb.2009.07.016 PMID:19660920

areGardea-Torresdey, J. L., Gomez, E., Peralta-Videa, J. R., Parsons, J. G., Troiani, H., & Jose-Yacaman, M. (2003). Alfalfa sprouts: A natural source for the synthesis of silver nanoparticles. *Langmuir*, *19*(4), 1357–1361. doi:10.1021/la020835i

Gardea-Torresdey, J. L., Parsons, J. G., Gomez, E., Peralta-Videa, J., Troiani, H. E., Santiago, P., & Yacaman, M. J. (2002). Formation and growth of Au nanoparticles inside live alfalfa plants. *Nano Letters*, *2*(4), 397–401. doi:10.1021/nl015673+

Gnanadesigan, M., Anand, M., Ravikumar, S., Maruthupandy, M., Vijayakumar, V., & Selvam, S. et al. (2011). Biosynthesis of silver nanoparticles by using mangrove plant extract and their potential mosquito larvicidal property. *Asian Pacific Journal of Tropical Medicine*, *4*(10), 799–803. doi:10.1016/S1995-7645(11)60197-1 PMID:22014736

Haefeli, C., Franklin, C., & Hardy, K. (1984). Plasmid-determined silver resistance in *Pseudomonas stutzeri* isolated from silver mine. *Journal of Bacteriology*, *158*, 389–392. PMID:6715284

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Hafner, J. H., Cheung, C. L., Woolley, A. T., & Lieber, C. M. (2001). Structural and functional imaging with carbon nanotube AFM probes. *Progress in Biophysics and Molecular Biology*, *77*(1), 73–110. doi:10.1016/S0079-6107(01)00011-6 PMID:11473787

Harisinghani, M. G., Barentsz, J., Hahn, P. F., Deserno, W. M., Tabatabaei, S., & van de Kaa, C. H. et al. (2003). Noninvasive detection of clinically occult lymph-node metastases in prostate cancer. *The New England Journal of Medicine*, *348*(25), 2491–2499. doi:10.1056/NEJMoa022749 PMID:12815134

Harris, A. T., & Bali, R. (2008). On the formation and extent of uptake of silver nanoparticles by live plants. *Journal of Nanoparticle Research*, *10*(4), 691–695. doi:10.1007/s11051-007-9288-5

He, S., Zhang, Y., Guo, Z., & Gu, N. (2008). Biological synthesis of gold nanowires using extract of Rhodopseudomonas capsulata. *Biotechnology Progress*, *24*(2), 476–480. doi:10.1021/bp0703174 PMID:18293997

Huang, J., Li, Q., Sun, D., Lu, Y., Su, Y., & Yang, X. et al. (2007). Biosynthesis of silver and gold nanopar- ticles by novel sundried Cinnamomum camphora leaf. *Nanotechnology*, *18*(10), 105104. doi:10.1088/0957-4484/18/10/105104

Husseiny, M. I., El-Aziz, M. A., Badr, Y., & Mahmoud, M. (2007). Biosynthesis of gold nanoparticles using Pseudomonas aeruginosa. *Spectrochimica acta. Part A, Molecular and biomolecular spectroscopy* or *Spectrochim Acta A Mol Biomol Spectrosc*, *67*(3-4), 1003–1006. doi:10.1016/j.saa.2006.09.028 PMID:17084659

Hütten, A., Sudfeld, D., Ennen, I., Reiss, G., Hachmann, W., & Heinzmann, U. et al. (2004). New magnetic nanoparticles for biotechnology. *Journal of Biotechnology*, *112*(1-2), 47–63. doi:10.1016/j.jbiotec.2004.04.019 PMID:15288940

Jain, D., Kachhwaha, S., Jain, R., Srivastava, G., & Hothari, S. L. (2010). Novel microbial route to synthesize silver nanoparticles using spore crystal mixture of Bacillus thuringiensis. *Indian Journal of Experimental Biology*, *48*, 1152–1156. PMID:21117457

Jain, N., Bhargava, A., Majumdar, S., Tarafdar, J. C., & Panwar, J. (2011). Extracellular biosynthesis and characterization of silver nanoparticles using Aspergillus flavus NJP08: A mechanism perspective. *Nanoscale*, *3*(2), 635–641. doi:10.1039/c0nr00656d PMID:21088776

Jayaseelan, C., Rahuman, A. A., Rajakumar, G., Santhoshkumar, T., Kirthi, A. V., & Marimuthu, S. et al. (2012). Efficacy of plant-mediated synthesized silver nanoparticles against hematophagous parasites. *Parasitology Research*, *111*(2), 921–933. doi:10.1007/s00436-011-2473-6 PMID:21638210

Jirage, K. B., Hulteen, J. C., & Martin, C. R. (1997). Nanotubule based molecular-filtration membranes. *Science*, *278*(5338), 655–658. doi:10.1126/science.278.5338.655

Joerger, R., Klaus, T., & Granqvist, C. G. (2000). Biologically produced silvercarbon composite materials for optically functional thin film coatings. *Advanced Materials*, *12*(6), 407–409. doi:10.1002/(SICI)1521-4095(200003)12:6<407::AID-ADMA407>3.0.CO;2-O

Kalimuthu, K., Babu, R.S., Venkataraman, D., Bilal, M. & Gurunathan, S. (2008). Biosynthesis of silver nanocrystals by Bacillus licheniformis. *Colloids Surfaces, B, 65,* 150–153.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Kalishwaralal, K., Deepak, V., Ram Kumar Pandian, S., & Gurunathan, S. (2009). Biological synthesis of gold nanocubes from Bacillus licheniformis. *Bioresource Technology*, *100*(21), 5356–5358. doi:10.1016/j.biortech.2009.05.051 PMID:19574037

Kalishwaralal, K., Deepak, V., Ramkumarpandian, S., Nellaiah, H., & Sangiliyandi, G. (2008). Extracellular biosynthesis of silver nanoparticles by the culture supernatant of Bacillus licheniformis. *Materials Letters*, *62*(29), 4411–4413. doi:10.1016/j.matlet.2008.06.051

Kalishwaralal, K., Gopalram, S., Vaidyanathan, R., Deepak, V., Pandian, S. R. K., & Gurunathan, S. (2010). Optimization of α-amylase production for the green synthesis of gold nanoparticles. *Colloids and surfaces. B, Biointerfaces*, *77*(2), 174–180. doi:10.1016/j.colsurfb.2010.01.018 PMID:20189782

Kathiresan, K., Manivannan, S., Nabeel, M. A., & Dhivya, B. (2009). Studies on silver nanoparticles synthesized by a marine fungus, Penicillium fellutanum isolated from coastal mangrove sediment. *Colloids and surfaces. B, Biointerfaces*, *71*(1), 133–137. doi:10.1016/j.colsurfb.2009.01.016 PMID:19269142

Kaviya, S., Santhanalakshmi, J., Viswanathan, B., Muthumary, J., & Srinivasan, K. (2011). Biosynthesis of silver nanoparticles using Citrus sinensis peel extract and its antibacterial activity. *Spectrochimica acta. Part A, Molecular and biomolecular spectroscopy*, *79*(3), 594–598. doi:10.1016/j.saa.2011.03.040 PMID:21536485

Klaus, T., Joerger, R., Olsson, E., & Granqvist, C.-G. (1999). Silver-based crystalline nanoparticles, microbially fabricated. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(24), 13611–13614. doi:10.1073/pnas.96.24.13611 PMID:10570120

Kneuer, C., Sameti, M., Bakowsky, U., Schiestel, T., Schirra, H., Schmidt, H., & Lehr, C. M. (2000). A nonviral DNA delivery system based on surface modified silica-nanoparticles can efficiently transfect cells in vitro. *Bioconjugate Chemistry*, *11*(6), 926–932. doi:10.1021/bc0000637 PMID:11087343

Konishi, Y., Ohno, K., Saitoh, N., Nomura, T., Nagamine, S., & Hishida, H. et al. (2007). Bioreductive deposition of platinum nanoparticles on the bacterium Shewanella algae. *Journal of Biotechnology*, *128*(3), 648–653. doi:10.1016/j.jbiotec.2006.11.014 PMID:17182148

Koper, O. B., Klabunde, J. S., Marchin, G. L., Klabunde, K. J., Stoimenov, P., & Bohra, L. (2002). Nanoscale powders and formulations with biocidal activity toward spores and vegetative cells of bacillus species, viruses, and toxins. *Current Microbiology*, *44*(1), 49–55. doi:10.1007/s00284-001-0073-x PMID:11727041

Kowshik, M., Ashtaputre, S., Kharrazi, S., Vogel, W., Urban, J., Kulkarni, S. K., & Paknikar, K. M. (2003). Extracellular synthesis of silver nanoparticles by a silver-tolerant yeast strain MKY3. *Nanotechnology*, *14*(1), 95–100. doi:10.1088/0957-4484/14/1/321

Kowshik, M., Deshmukh, N., Vogel, W., Urban, J., Kulkarni, S. K., & Paknikar, K. M. (2002b). Microbial synthesis of semiconductor CdS nanoparticles, their characterization, and their use in fabrication of an ideal diode. *Biotechnology and Bioengineering*, *78*(5), 583–588. doi:10.1002/bit.10233 PMID:12115128

Kowshik, M., Vogel, W., Urban, J., Kulkarni, S. K., & Paknikar, K. M. (2002a). Microbial synthesis of semiconductor PbS nanocrystallites. *Advanced Materials*, *14*(11), 815–818. doi:10.1002/1521-4095(20020605)14:11<815::AID-ADMA815>3.0.CO;2-K

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Krumov, N., Oder, S., Perner-Nochta, I., Angelov, A., & Posten, C. (2007). Accumulation of CdS nanoparticles by yeasts in a fed-batch bioprocess. *Journal of Biotechnology*, *132*(4), 481–486. doi:10.1016/j.jbiotec.2007.08.016 PMID:17900736

Królikowska, A., Kudelski, A., Michota, A., & Bukowska, J. (2003). SERS studies on the structure of thioglycolic acid monolayers on silver and gold. *Surface Science*, *532-535*, 227–232. doi:10.1016/S0039-6028(03)00094-3

Kumar, A., Mandal, S., Pasricha, R., Mandale, A. B., & Sastry, M. (2003). Investigation into the interaction between surface-bound alkyl amines and gold nanoparticles. *Langmuir*, *19*(15), 6277–6282. doi:10.1021/la034209c

Kumar, R., Roopan, S. M., Prabhakarn, A., Khanna, V. G., & Chakroborty, S. (2012). waste Annona squamosa peel extract: Biosynthesis of silver nanoparticles. *Spectrochimica acta. Part A, Molecular and biomolecular spectroscopy*, *90*, 173–176. doi:10.1016/j.saa.2012.01.029 PMID:22336049

Labrenz, M., Druschel, G. K., Thomsen, E. T., Gilbert, B., Welch, S. A., & Kemner, K. M. et al. (2000). Formation of sphalerite (ZnS) deposits in natural biofilms of sulfate-reducing bacteria. *Science*, *290*(5497), 1744–1747. doi:10.1126/science.290.5497.1744 PMID:11099408

Lee, J.-H., Han, J., Choi, H., & Hur, H.-G. (2007). Effects of temperature and dissolved oxygen on Se(IV) removal and Se(0) precipitation by Shewanella sp. *HN-41. Chemosphere*, *68*(10), 1898–1905. doi:10.1016/j.chemosphere.2007.02.062 PMID:17434567

Lengke, M. F., Fleet, M. E., & Southam, G. (2006a). Morphology of gold nanoparticles synthesized by filamentous cyanobacteria from gold(I)-thiosulfate and gold(III)- chloride complexes. *Langmuir*, *22*(6), 2780–2787. doi:10.1021/la052652c PMID:16519482

Li, G., He, D., Qian, Y., Guan, B., Gao, S., & Cui, Y. et al. (2012). Fungus- mediated green synthesis of silver nanoparticles using Aspergillus terreus. *International Journal of Molecular Sciences*, *13*(12), 466–476. doi:10.3390/ijms13010466 PMID:22312264

Li, Y., Leung, P., Yao, L., Song, Q. W., & Newton, E. (2006). Antimicrobial effect of surgical masks coated with nanoparticles. *The Journal of Hospital Infection*, *62*(1), 58–63. doi:10.1016/j.jhin.2005.04.015 PMID:16099072

Lovley, D. R., Stolz, J. F., Nord, G. L. Jr, & Phillips, E. J. P. (1987). Anaerobic production of magnetite by a dissimilatory iron-reducing microorganism. *Nature*, *330*(6145), 252–254. doi:10.1038/330252a0

Lovley, D. R., Stolz, J. F., Nord, G. L. Jr, & Phillips, E. J. P. (1987). Anaerobic production of magnetite by a dissimilatory iron-reducing microorganism. *Nature*, *330*(6145), 252–254. doi:10.1038/330252a0

Lukman, A. I., Gong, B., Marjo, C. E., Roessner, U., & Harris, A. T. (2011). Facile synthesis, stabilization, and anti-bacterial performance of discrete Ag nanoparticles using Medicago sativa seed exudates. *Journal of Colloid and Interface Science*, *353*(2), 433–444. doi:10.1016/j.jcis.2010.09.088 PMID:20974473

Mandal, D., Bolander, M. E., Mukhopadhyay, D., Sarkar, G., & Mukherjee, P. (2006). The use of microorganisms for the formation of metal nanoparticles and their application. *Applied Microbiology and Biotechnology*, *69*(5), 485–492. doi:10.1007/s00253-005-0179-3 PMID:16317546

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Marimuthu, S., Rahuman, A. A., Rajakumar, G., Santhoshkumar, T., Kirthi, A. V., & Jayaseelan, C. et al. (2011). Evaluation of green synthesized silver nanoparticles against parasites. *Parasitology Research*, *108*(6), 1541–1549. doi:10.1007/s00436-010-2212-4 PMID:21181192

Mohanpuria, P., Rana, K. N., & Yadav, S. K. (2008). Biosynthesis of nanoparticles: Technological concepts and future applications. *Journal of Nanoparticle Research*, *10*(3), 507–517. doi:10.1007/s11051-007-9275-x

Pető, G., Molnar, G. L., Paszti, Z., Geszti, O., Beck, A., & Guczi, L. (2002). Electronic structure of gold nanoparticles deposited on SiOx/Si. *Materials Science and Engineering C*, *19*(1-2), 95–99. doi:10.1016/S0928-4931(01)00449-0

Moon, J.-W., Roh, Y., Lauf, R. J., Vali, H., Yeary, L. W., & Phelps, T. J. (2007). Microbial preparation of metal-substituted magnetite nanoparticles. *Journal of Microbiological Methods*, *70*(1), 150–158. doi:10.1016/j.mimet.2007.04.012 PMID:17532071

Mukherjee, P., Ahmad, A., Mandal, D., Senapati, S., Sainkar, S. R., & Khan, M. I. et al. (2001). Fungus mediated synthesis of silver nanoparticles and their immobilization in the mycelial matrix: A novel biological approach to nanoparticle synthesis. *Nano Letters*, *1*(10), 515–519. doi:10.1021/nl0155274

Mukherjee, P., Roy, M., Mandal, B. P., Dey, G. K., Mukherjee, P. K., & Ghatak, J. et al. (2008). Green synthesis of highly stabilized nanocrystalline silver particles by a non-pathogenic and agriculturally important fungus T. asperellum. *Nanotechnology*, *19*(7), 075103. doi:10.1088/0957-4484/19/7/075103 PMID:21817628

Mukherjee, P., Senapati, S., Mandal, D., Ahmad, A., Khan, M. I., Kumar, R., & Sastry, M. (2002). Extracellular synthesis of gold nanoparticles by the fungus Fusarium oxysporum. *ChemBioChem*, *3*(5), 461–463. doi:10.1002/1439-7633(20020503)3:5<461::AID-CBIC461>3.0.CO;2-X PMID:12007181

Mullen, M. D., Wolf, D. C., Ferris, F. G., Beveridge, T. J., Flemming, C. A., & Bailey, G. W. (1989). Bacterial sorption of heavy metals. *Applied and Environmental Microbiology*, *55*, 3143–3149. PMID:2515800

Nair, B., & Pradeep, T. (2002). Coalescence of nanoclusters and formation of submicron crystallites assisted by Lactobacillus strains. *Crystal Growth & Design*, *2*(4), 293–298. doi:10.1021/cg0255164

Nanda,, A., & Shahnaz, M. (2014). Enhanced antibacterial efficacy of biosynthesized AgNPs from Penicillium glabrum (MTCC1985) pooled with different drugs. *International Journal of Pharm Tech Research*, *6*(1), 217–223.

Nanda, A., & Saravanan, M. (2009). Biosynthesis of silver nanoparticles from Staphy- lococcus aureus and its antimicrobial activity against MRSA and MRSE. *Nanomedicine; Nanotechnology, Biology, and Medicine*, *5*(4), 452–456. doi:10.1016/j.nano.2009.01.012 PMID:19523420

Narayanan, K. B., & Sakthivel, N. (2008). Coriander leaf mediated biosynthesis of gold nanoparticles. *Materials Letters*, *62*(30), 4588–4590. doi:10.1016/j.matlet.2008.08.044

Nayak B. K., Bhat, M. A., & Nanda, A. (2014). Amplification of antibacterial property of drugs in combination with silver nanoparticles synthesized from an airborne fungus. *International Journal of Chemtech Research*, *6* (3).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Nayak, B. K., & Bhat, M., A., & Nanda, A. (2013). Implementation of *penicillium sp*. as raw material for synthesizing metal nanoparticles for antibiosis. *Functional Nanomaterials for Energy and Environmental Applications*, *760*(1), 33–38.

Nolasco-Arismendi, V., Morales-Luckie, R., Sánchez-Mendieta, V., Hinestroza, J. P., Castro-Longoria, E., & Vilchis-Nestor, A. R. (2012). Formation of silk–gold nanocomposite fabric using grapefruit aqueous extract. *Textile Research Journal*, 1–7.

Panda, K. K., Achary, V. M., Krishnaveni, R., Padhi, B. K., Sarangi, S. N., Sahu, S. N., & Panda, B. B. (2011). In vitro biosynthesis and genotoxicity bioassay of silver nanoparticles using plants. *Toxicology In Vitro*, *25*(5), 1097–1105. doi:10.1016/j.tiv.2011.03.008 PMID:21419840

Parashar, U. K., Saxena, S. P., & Srivastava, A. (2009). Bioinspired synthesis of silver nanoparticles. *Digest Journal of Nanomaterials and Biostructures*, *4*, 159–166.

Parikh, R. Y., Singh, S., Prasad, B. L., Patole, M. S., Sastry, M., & Shouche, Y. S. (2008). Extracellular synthesis of crystalline silver nanoparticles and molecular evidence of silver resistance from Morganella. sp.: Towards understanding biochemical synthesis mechanism. *ChemBioChem*, *9*(9), 1415–1422. doi:10.1002/cbic.200700592 PMID:18491326

Philip, D. (2009). Biosynthesis of Au, Ag and Au–Ag nanoparticles using edible mush- room extract. *Spectrochimica acta. Part A, Molecular and biomolecular spectroscopy*, *73*(2), 374–381. doi:10.1016/j.saa.2009.02.037 PMID:19324587

Philip, D. (2010). Green synthesis of gold and silver nanoparticles using Hibiscus rosa sinensis. *Physica E, Low-Dimensional Systems and Nanostructures*, *42*(5), 1417–1424. doi:10.1016/j.physe.2009.11.081

Philipse, A. P., & Maas, D. (2002). Magnetic colloids from magnetotactic bacteria: Chain formation and colloidal stability. *Langmuir*, *18*(25), 9977–9984. doi:10.1021/la0205811

Pollmann, K., Merroun, M., Raff, J., Hennig, C., & Selenska-Pobell, S. (2006). Manufacturing and characterization of Pd nanoparticles formed on immobilized bacterial cells. *Letters in Applied Microbiology*, *43*(1), 39–45. doi:10.1111/j.1472-765X.2006.01919.x PMID:16834719

Prakasham, R. S., Buddana, S. K., Yannam, S. K., & Guntuku, G. S. (2012). Characterization of silver nanoparticles synthesized by using marine isolate Streptomyces albido flavus. *Journal of Microbiology and Biotechnology*, *22*(5), 614–621. doi:10.4014/jmb.1107.07013 PMID:22561854

Rajakumar, G., Rahuman, A. A., Roopan, S. M., Khanna, V. G., Elango, G., & Kamaraj, C. et al. (2012). Fungus-mediated biosynthesis an characterization of TiO2 nanoparticles and their activity against *pathogenic bacteria. Spectrochimica acta. Part A, Molecular and biomolecular spectroscopy*, *91*, 23–29. doi:10.1016/j.saa.2012.01.011 PMID:22349888

Reddy, A. S., Chen, C. Y., Chen, C. C., Jean, J. S., Chen, H. R., & Tseng, M. J. et al. (2010). Biological synthesis of gold and silver nanoparticles mediated by the bacteria Bacillus subtilis. *Journal of Nanoscience and Nanotechnology*, *10*(10), 6567–6574. doi:10.1166/jnn.2010.2519 PMID:21137763

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Renehan, E. M., Enneking, F. K., Varshney, M., Partch, R., Dennis, D. M., & Morey, T. E. (2005). Scavenging nanoparticles: An emerging treatment for local anesthetic toxicity. *Regional Anesthesia and Pain Medicine*, *30*, 380–384. PMID:16032590

Robinson, M. G., Brown, L. N., & Hall, B. D. (1997). Effect of Gold (III) on the Pauling Diatom *Amphora ceffeaeformis* uptake, toxicity and interactions with copper. *Biofouling*, *11*(1), 59–79. doi:10.1080/08927019709378320

Roh, Y., Jang, H. D., & Suh, Y. (2007). Microbial synthesis of magnetite and Mn-substituted magnetite nanoparticles: *Influence of bacteria and incubation temperature. Journal of Nanoscience and Nanotechnology*, *7*(11), 3938–3943. doi:10.1166/jnn.2007.076 PMID:18047092

Roopan, S. M., Bharathi, A., Prabhakarn, A., Abdul Rahuman, A., Velayutham, K., & Rajakumar, G. et al. (2012). Efficient phyto-synthesis and structural characterization of rutile TiO2 nanoparticles using Annona squamosa peel extract. *Spectrochim*, *98*, 86–90. doi:10.1016/j.saa.2012.08.055

Rosi, N. L., & Mirkin, C. A. (2005). Nanostructures in biodiagnostics. *Chemical Reviews*, *105*(4), 1547–1562. doi:10.1021/cr030067f PMID:15826019

Sadhasivam, S., Shanmugam, P., Veerapandian, M., Subbiah, R., & Yun, K. (2012). Biogenic synthesis of multidimensional gold nanoparticles assisted by Streptomyces hygroscopicus and its electrochemical and antibacterial properties. *Biometals*, *25*(2), 351–360. doi:10.1007/s10534-011-9506-6 PMID:22069027

Sadhasivam, S., Shanmugam, P., & Yun, K. S. (2010). Biosynthesis of silver nanoparticles by Streptomyces hygroscopicus and antimicrobial activity against medically important pathogenic microorganisms. *Colloids and surfaces. B, Biointerfaces*, *81*(1), 358–362. doi:10.1016/j.colsurfb.2010.07.036 PMID:20705438

Samadi, N., Golkaran, D., Eslamifar, A., Jamalifar, H., Fazeli, M. R., & Mohseni, F. A. (2009). Intra/extracellular biosynthesis of silver nanoparticles by an autochthonous strain of Proteus mirabilis isolated from photographic waste. *Journal of Biomedical Nanotechnology*, *5*(3), 247–253. doi:10.1166/jbn.2009.1029 PMID:20055006

Saravanan, M., & Nanda, A. (2010). Extracellular synthesis of silver bionanoparticles from Aspergillus clavatus and its antimicrobial activity against MRSA and *MRSE. Colloids and surfaces. B, Biointerfaces*, *77*(2), 214–218. doi:10.1016/j.colsurfb.2010.01.026 PMID:20189360

Sarkar, J., Ray, S., Chattopadhyay, D., Laskar, A., & Acharya, K. (2012). Mycogenesis of gold nanoparticles using a phytopathogen *Alternaria alternata. Bioprocess and Biosystems Engineering*, *35*(4), 637–643. doi:10.1007/s00449-011-0646-4 PMID:22009439

Sathishkumar, M., Sneha, K., Won, S. W., Cho, C. W., Kim, S., & Yun, Y. S. (2009). Cinnamon zeylanicum bark extract and powder mediated green synthesis of nano-crystalline silver particles and its bacterial activity. *Colloids and surfaces. B, Biointerfaces*, *73*(2), 332–338. doi:10.1016/j.colsurfb.2009.06.005 PMID:19576733

Satyavani, K., Gurudeeban, S., Ramanathan, T., & Balasubramanian, T. (2011). Biomedical potential of silver nanoparticles synthesized from calli cells of Citrullus colocynthis (L.) Schrad. *Journal of Nanobiotechnology*, 43. PMID:21943321

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Shankar, S. S., Ahmad, A., Pasricha, R., Khan, M. I., Kumar, R., & Sastry, M. (2004a). Immobilization of biogenic gold nanoparticles in thermally evaporated fatty acid and amine thin films. *Journal of Colloid and Interface Science*, *274*(1), 69–75. doi:10.1016/j.jcis.2003.12.011 PMID:15120279

Shankar, S. S., Ahmad, A., Pasricha, R., & Sastry, M. (2003a). Bioreduction of chloroaurate ions by geranium leaves and its endophytic fungus yields gold nanoparticles of different shapes. *Journal of Materials Chemistry*, *13*(7), 1822–1826. doi:10.1039/b303808b

Shankar, S. S., Ahmad, A., & Sastry, M. (2003b). Geranium leaf assisted biosynthesis of silver nanoparticles. *Biotechnology Progress*, *19*(6), 1627–1631. doi:10.1021/bp034070w PMID:14656132

Shankar, S. S., Rai, A., Ahmad, A., & Sastry, M. (2004b). Rapid synthesis of Au, Ag, and bimetallic Au core–Ag shell nanoparticles using Neem (Azadirachta indica) leaf broth. *Journal of Colloid and Interface Science*, *275*(2), 496–502. doi:10.1016/j.jcis.2004.03.003 PMID:15178278

Sintubin, L., Windt, W., Dick, J., Mast, J., Ha, D., Verstraete, W., & Boon, N. (2009). Lactic acid bacteria as reducing and capping agent for the fast and efficient production of silver nanoparticles. *Applied Microbiology and Biotechnology*, *84*(4), 741–749. doi:10.1007/s00253-009-2032-6 PMID:19488750

Smith, P. R., Holmes, J. D., Richardson, D. J., Russell, D. A., & Sodeau, J. R. (1998). Photophysical and photochemical characterization of bacterial semiconductor cadmium sulphide particles. *Journal of the Chemical Society*, *94*, 1235–1241.

Sondi, I., & Salopek-Sondi, B. (2004). Silver nanoparticles as antimicrobial agent: A case study of E. coli as a model for Gram-negative bacteria. *Journal of Colloid and Interface Science*, 177. PMID:15158396

Song, J. Y., Jang, H.-K., & Kim, B. S. (2009). Biological synthesis of gold nanoparticles using Magnolia kobus and Diopyros kaki leaf extracts. *Process Biochemistry*, *44*(10), 1133–1138. doi:10.1016/j.procbio.2009.06.005

Southam, G., & Beveridge, T. J. (1996, November). The occurrence of sulfur and phosphorus within bacterially derived crystalline and pseudocrystalline octahedral gold formed in vitro. *Geochimica et Cosmochimica Acta*, *60*(22), 4369–4376. doi:10.1016/S0016-7037(96)00235-9

Spring, S., & Schleifer, K. H. (1995). Diversity of magnetotactic bacteria. *Systematic and Applied Microbiology*, *18*(2), 147–153. doi:10.1016/S0723-2020(11)80386-3

Suresh, A. K., Pelletier, D. A., Wang, W., Broich, M. L., Moon, J. W., & Gu, B. et al. (2011). Biofabrication of discrete spherical gold nanoparticles using the metal-reducing bacterium Shewanella oneidensis. *Acta Biomaterialia*, *7*(5), 2148–2152. doi:10.1016/j.actbio.2011.01.023 PMID:21241833

Taniguchi, N. (1974). On the basic concept of 'Nano-Technology'. In *Proceedings of International Conference on Production Engineering*. Japan Society of Precision Engineering.

Uhrich, K. E., Cannizzaro, S. M., Langer, R. S., & Shakesheff, K. M. (1999). Polymeric systems for controlled drug release. *Chemical Reviews*, *99*(11), 3181–3198. doi:10.1021/cr940351u PMID:11749514

Vaidyanathan, R., Gopalram, S., Kalishwaralal, K., Deepak, V., Pandian, S. R. K., & Gurunathan, S. (2010). Enhanced silver nanoparticle synthesis by optimization of nitrate reductase activity. *Colloids Surfaces B*, *75*(1), 335–341. doi:10.1016/j.colsurfb.2009.09.006 PMID:19796922

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Velayutham, K., Rahuman, A. A., Rajakumar, G., Roopan, S. M., Elango, G., & Kamaraj, C. et al. (2013). Larvicidal activity of green synthesized silver nanoparticles using bark aqueous extract of Ficus racemosa against Culex quinquefasciatus and Culex gelidus. *Asian Pacific Journal of Tropical Medicine*, *6*(2), 95–101. doi:10.1016/S1995-7645(13)60002-4 PMID:23339909

Velmurugan, P., Shim, J., You, Y., Choi, S., Kamala-Kannan, S., & Lee, K.-J. et al. (2010). Removal of zinc by live, dead, and dried biomass of Fusarium spp. isolated from the abandoned-metal mine in South Korea and its perspective of producing nanocrystals. *Journal of Hazardous Materials*, *182*(1-3), 317–324. doi:10.1016/j.jhazmat.2010.06.032 PMID:20599320

Vigneshwaran, N., Ashtaputre, N. M., Varadarajan, P. V., Nachane, R. P., Paralikar, K. M., & Balasubramanya, R. H. (2007b). Biological synthesis of silver nanoparticles using the fungus Aspergillus flavus. *Materials Letters*, *61*(6), 1413–1418. doi:10.1016/j.matlet.2006.07.042

Vigneshwaran, N., Kathe, A. A., Varadarajan, P. V., Nachane, R. P., & Balasubramanya, R. H. (2006). Biomimetics of silver nanoparticles by white rot fungus, Phaenerochete chrysosporium. *Colloids and surfaces. B, Biointerfaces*, *53*(1), 55–59. doi:10.1016/j.colsurfb.2006.07.014 PMID:16962745

Vigneshwaran, N., Kathe, A. A., Varadarajan, P. V., Nachane, R. P., & Balasubramanya, R. H. (2007a). Silver–protein (Core–Shell) nanoparticle production using spent mushroom substrate. *Langmuir*, *23*(13), 7113–7117. doi:10.1021/la063627p PMID:17518485

Watson, J. H. P., Cressey, B. A., Roberts, A. P., Ellwood, D. C., Charnock, J. M., & Sopet, A. K. (2000). Structural and magnetic studies on heavy-metal-adsorbing iron sulphide nanoparticles produced by sulphate-reducing bacteria. *Journal of Magnetism and Magnetic Materials*, *214*(1), 3–30.

Watson, J. H. P., Ellwood, D. C., Soper, A. K., & Charnock, J. (1999). Nanosized strongly magnetic bacterially-produced iron sulfide materials. *Journal of Magnetism and Magnetic Materials*, *203*(1-3), 69–72. doi:10.1016/S0304-8853(99)00191-2

Weinberg, G., Ripper, R., Feinstein, D. L., & Hoffman, W. (2003). Lipid emulsion infusion rescues dogs from bupivacaine-induc cardiac toxicity. *Regional Anesthesia and Pain Medicine*, *28*(3), 198–202. doi:10.1097/00115550-200305000-00005 PMID:12772136

Weissleder, R., Lewin, M., Carlesso, N., Tung, C.-H., Tang, X.-W., Cory, D., & Scadden, D. T. (2000). Tat peptide derivatized magnetic nanoparticles allow in vivo tracking and recovery of progenitor cells. *Nature Biotechnology*, *18*(4), 410–414. doi:10.1038/74464 PMID:10748521

Wörner, M., Lioubashevski, O., Basel, M. T., Niebler, S., Gogritchiani, E., & Egner, N. et al. (2007). Characterization of nanostructured surfaces generated by reconstitution of the porin Msp from Mycobacterium smegmatis. *Small*, *3*(6), 1084–1097. doi:10.1002/smll.200600559 PMID:17514767

Xie, J., Lee, J. Y., Wang, D. I. C., & Ting, Y. P. (2007). High-yield synthesis of complex gold nanostructures in a fungal system. *Journal of Physical Chemistry*, *C*, 111, 16858–16865.

Zhang, H., Li, Q., Lu, Y., Sun, D., Lin, X., & Deng, X. et al. (2005). Biosorption and bioreduction of diamine silver complex by Corynebacterium. *Journal of Chemical Technology and Biotechnology (Oxford, Oxfordshire)*, *80*(3), 285–290. doi:10.1002/jctb.1191

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Zhang, H., Li, Q., Wang, H., Sun, D., Lu, Y., & He, N. (2007). Accumulation of silver (I) ion and diamine silver complex by Aeromonas SH10 biomass. *Applied Biochemistry and Biotechnology*, *143*(1), 54–62. doi:10.1007/s12010-007-8006-1 PMID:18025596

ADDITIONAL READING

Ahmad, A., Mukherjee, P., Senapati, S., Mandal, D., Khan, M. I., Kumar, R., & Sastry, M. (2003). Extracellular biosynthesis of silver nanoparticles using the fungus Fusarium oxysporum. *Colloids and Surfaces. B, Biointerfaces*, *28*(4), 313–318. doi:10.1016/S0927-7765(02)00174-1

Durán, N., Marcato, P.D., Alves, O.L., De Souza, G.I.H., & Esposito, E., (2005). Mechanistic aspects of biosynthesis of silver nanoparticles by several Fusarium oxysporum strains. *Journal of Nanobiotechnol.3-8*

Kalishwaralal, K., Deepak, V., Ramkumarpandian, S., Nellaiah, H., & Sangiliyandi, G. (2008). Extracellular biosynthesis of silver nanoparticles by the culture supernatant of Bacillus licheniformis. *Materials Letters*, *62*(29), 4411–4413. doi:10.1016/j.matlet.2008.06.051

Mohanpuria, P., Rana, K. N., & Yadav, S. K. (2008). Biosynthesis of nanoparticles: Technological concepts and future applications. *Journal of Nanoparticle Research*, *10*(3), 507–517. doi:10.1007/s11051-007-9275-x

Mukherjee, P., Ahmad, A., Mandal, D., Senapati, S., Sainkar, S. R., & Khan, M. I. et al. (2001). Fungus mediated synthesis of silver nanoparticles and their immobilization in the mycelial matrix: A novel biological approach to nanoparticle synthesis. *Nano Letters*, *1*(10), 515–519. doi:10.1021/nl0155274

Mukherjee, P., Roy, M., Mandal, B. P., Dey, G. K., Mukherjee, P. K., & Ghatak, J. et al. (2008). Green synthesis of highly stabilized nanocrystalline silver particles by a non-pathogenic and agriculturally important fungus T. asperellum. *Nanotechnology*, *19*(7), 075103. doi:10.1088/0957-4484/19/7/075103 PMID:21817628

Nanda, A. & Shahnaz, M. (2013). Comparative analysis of antibiotic effect of $AgNO_3$ and silver nanoparticles synthesized from *Penicillium fellutanum. Functional Nanomaterials for Energy and Environmental Applications*, *760*(1), 9–14.

Nanda, A., & Saravanan, M. (2009). Biosynthesis of silver nanoparticles from Staphy- lococcus aureus and its antimicrobial activity against MRSA and MRSE. *Nanomedicine; Nanotechnology, Biology, and Medicine*, *5*(4), 452–456. doi:10.1016/j.nano.2009.01.012 PMID:19523420

Nayak, B. K., & Bhat, M. A., & Nanda, A. (2013). Implementation of *Penicillium sp*. As raw material for Synthesizing Metal Nanoparticles for Antibiosis. *Functional Nanomaterials for Energy and Environmental Applications*, *760*(1), 33–38.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

KEY TERMS AND DEFINITIONS

Bioremediation: It is a waste management technique that involves the use of organisms to remove or neutralize pollutants from a contaminated site.

Carbon Nanotubes (CNTs): Hollow, cylindrical structures, with a diameter usually less than 5 nanometres. They are often but not necessarily, composed of carbon, and having remarkable strength and electrical properties.

Liposome: Phospholipid bilayer vesicle formed from an aqueous suspension of phospholipid molecules.

Nanomaterials: A broad class of materials that includes nanometre-sized crystallites, nanometre, sized powders, macroscopic objects, and films composed of nanometre sized particles or any phases combined with a spatial distribution that involves nanometre length scale of less than 100 nm in at least one dimension.

Nanomedicine: The use of nanotechnology in the improvement of a contaminated site to prevent, minimize or mitigate damage to human health or the environment.

Nanoparticles: The term refers to particles of sizes in the range of 1–100 nm. Depending on the preparative conditions, nanoparticles may have polycrystalline or single crystalline structure. Nanoparticles of sizes less than 5 nm are termed as quantum dots.

Nanosensor: A device for sensing radiation, forces, chemicals, or biological agents, in which some portion of a device operates at the nanoscale.

Nanotechnology: A manufacturing technology able to inexpensively fabricate most structures consistent with natural law, and to do so with molecular precision.

Quantum Dots: Nonocrystals or nanoparticles. Refer to confined electrons. Electrons in them occupy discrete states as in the case of atoms and therefore, quantum dots are referred to as artificial atoms.

Remediation: Remediation is the science of removal or reduction of pollutants from the environment using chemical or biological means.

This work was previously published in Nanotechnology Applications for Improvements in Energy Efficiency and Environmental Management edited by M. A. Shah, M. A. Bhat, and J. Paulo Davim, pages 116-155, copyright year 2015 by Information Science Reference (an imprint of IGI Global).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

APPENDIX

- Nanotechnology is technology on the atomic and molecular scale.
- A nanometre (nm) is one billionth of a meter.
- A nanoparticle is a particle with one or more external dimensions in the size range 1-100 nm.
- The aspect ratio between a nanoparticle and a football is similar to that between a football and Earth.
- Nanotechnology is working on a scale of 100 nm (which corresponds approximately to the size of a virus) down to the size of atoms, about 0.1 nm.
- Nano-scale materials and processes are present in nature, ranging from free molecules in gases and liquids to proteins and organic processes.
- Some substances are produced unintentionally, such as welding dust and diesel particulates.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 11
Diverse Applications of Nanotechnology in Biomedicine, Chemistry, and Engineering

Sivapriya Kirubakaran
Indian Institute of Technology Gandhinagar, India

Vijay Thiruvenkatam
Indian Institute of Technology Gandhinagar, India

ABSTRACT

This chapter introduces the basic concepts of nano science to readers. Some novel methodologies for synthesizing nano particles are discussed briefly. Since the book title suggests diverse applications of nano materials, this chapter also summarizes the applications of nano technology in medicine (nano medicine), where tissue engineering and regenerative medicine are discussed. Other applications DNA nanotechnology in living organisms are discussed briefly. Overall, this chapter introduces the readers to broad sections of nano science and its applications in chemistry, engineering, and medicine.

HISTORY

"There is plenty of room at the bottom" a famous quote by Richard Feynman which was inspirational to a new field of science called nanotechnology. Later in 1980 Eric Drexler proposed the idea of a nanoscale assembler, which can build a copy of itself and also of other items and finally become a complex system. Invention of scanning tunneling microscope led to the discovery of fullerens in 1985 and carbon nanotubes followed by it.

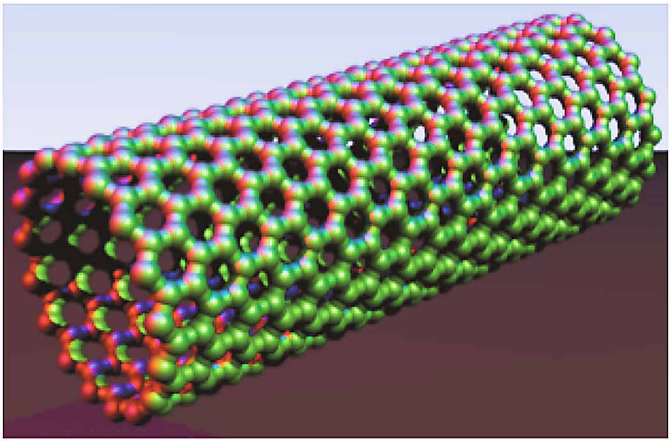
Nanotubes are the first well studied nano materials and belong to fullerene structural family. The name is derived from their long, hollow tube like structure formed by one atom thick sheets of carbon. These sheets are also rolled at specific chiral angles and these properties decide the nature of the nanotubes. (Figure 1) Coarse particles cover a range between 2,500 and 10,000 nanometers, fine particles are sized between 100 and 2,500 nanometers. Ultrafine particles, or nanoparticles, are between 1 and 100 nanometers in size. Nanoparticles are of great interest to scientific community as they form a bridge

DOI: 10.4018/978-1-5225-0549-5.ch011

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 1. Carbon nanotube
Credit: Wiki:http://en.wikipedia.org/wiki/Carbon_nanotube



between bulk materials and atomic structures. A bulk material should have constant physical properties regardless of its size, but at the nano-scale size-dependent properties are often observed. Thus, the properties of materials change as their size approaches the nanoscale and as the percentage of atoms at the surface of a material becomes significant. The interesting and sometimes unexpected properties of nanoparticles are therefore largely due to the large surface area of the material, which dominates the contributions made by the small bulk of the material.

SYNTHESIS

There are several methods reported to synthesis variety of nano materials. Depend on their application the synthetic method varies. Though the synthesis and organization of nanoparticles provide complementary tools for nanotechnology, processing of nanoparticles into bulk shapes while retaining their nanosize is a challenge as far as structural and engineering applications are concerned. Synthesis and assembly strategies of nanoparticles mostly accommodate precursors from liquid, solid or gas phase; employ chemical or physical deposition approaches; and similarly rely on either chemical reactivity or physical compaction to integrate the nanostructure building blocks within the final material structure. The variety of techniques that can be classified in top-down or bottom-up approaches. (Mijatovic, Eijkel, & van den Berg, 2005)

The bottom- up approach of nanomaterials synthesis first forms the nanostructured building blocks (nanoparticles). Build-up of a material from the bottom: atom-by-atom molecule-by-molecule. In organic chemistry and/or polymer science, we know polymers are synthesized by connecting individual monomers together. In crystal growth, growth species, such as atoms, ions and molecules, after impinging onto the growth surface, assemble into crystal structure one after another. Examples: Production of salt and nitrate in chemical industry, Growth of single crystals and deposition of films in electronic industry. For most materials, there is no difference in physical properties of materials regardless of the synthesis routes, provided that chemical composition, crystallinity, and microstructure of the material in question are identical. (Wegner et al., 2004)

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

In top down approaches, the source material is reduced from bulk size to nanoscale scale as in attrition processes (e.g., grinding). Grinding is a mechanical attrition process, which operates on the solid phase. (Heidenreich, Büttner & Ebert, 2003) An example of this approach is the formation of powder components through aerosol and sol-gel techniques and then the compaction of the components into the final material. Nanoparticles with diameters ranging from 1 nm to 10 nm with consistent crystal structure, surface derivatization, and a high degree of monodispersity can be processed by these techniques. (Lenggoro,Okuyama, de la Mora & Tohge, 2000). Gas phase synthesis or aerosol processing of nanoparticles is based on evaporation and condensation (nucleation and growth) in a sub atmospheric inert- gas environment; and the processing mainly is by combustion flame, laser ablation, chemical vapour condensation, spray pyrolysis, electro spray etc. (Iskandar, 2009) On the other hand, sol-gel processing is a wet chemical synthesis approach that can be used to generate nanoparticles by gelation, precipitation, and hydrothermal treatment.

APPLICATIONS

Biomedicine (Nanomedicine)

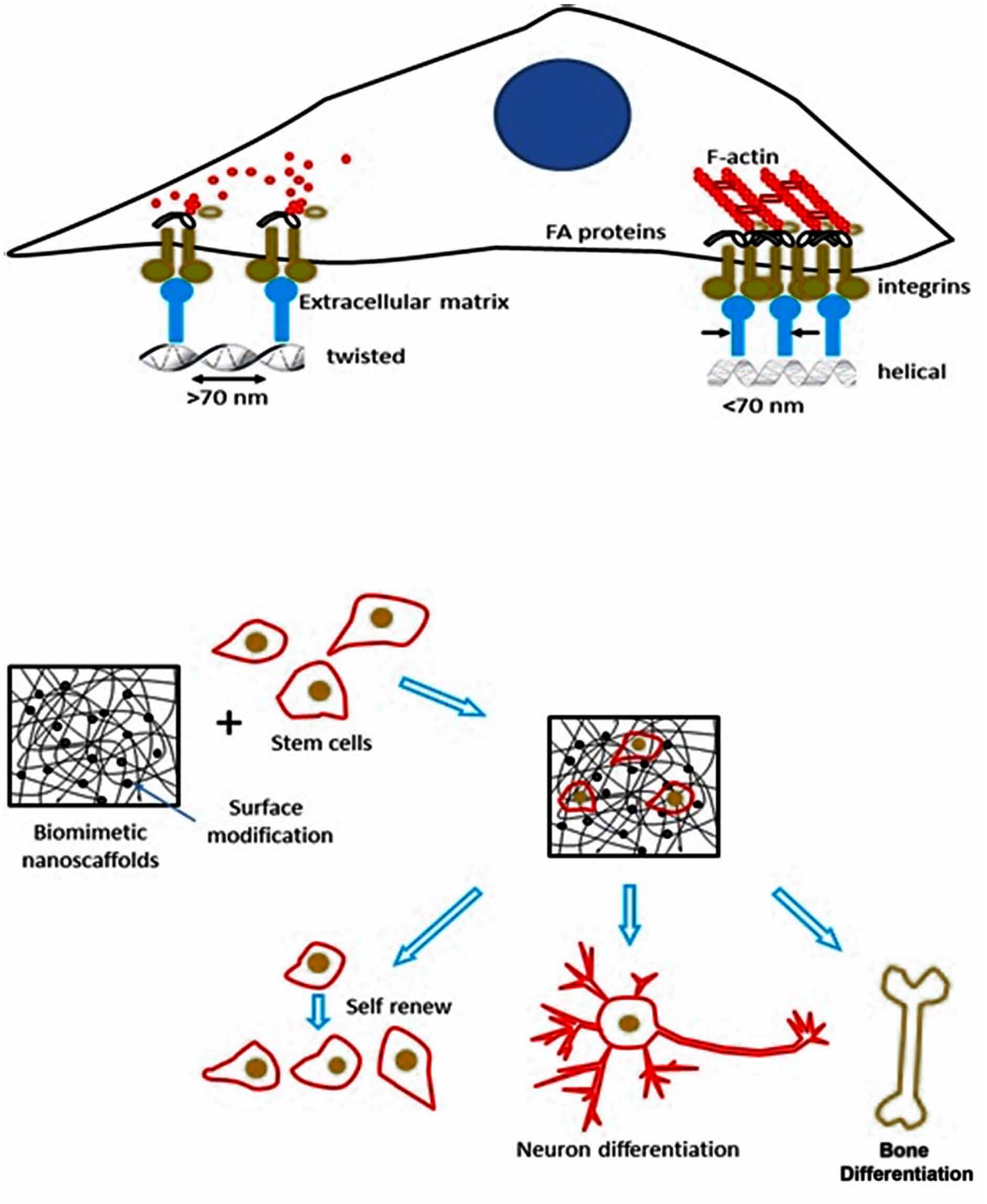
Significant interest has arisen in the application of nanotechnology in medicine and biomedical applications. The integration of nanotechnology into the field of medical science has opened new possibilities. Nanomaterials has allowed a better understanding of molecular biology. As a consequence, there is the potential of providing novel methods for the treatment of diseases which were previously difficult to target due to size restrictions.

Here we summarize a few of the studies that were path-breaking studies in biomedicine. One of the recent achievements is reducing the effect of radiation by using gold nanorod assisted thermal therapy. (Soni, Tyagi, Taylor & Kumar 2013) This is to achieve the thermal ablation temperature (50–60°C) in the tumour region and spare healthy tissue surrounding the tumour. These studies have already been tested in mice and are awaiting human trials that will be using gold nanoshells to help diagnose and treat cancer. (Zhongshi, Li, Xie, and Liu 2014) The advantage of using nanoscale for medical technologies is that smaller devices are less invasive, can possibly be implanted inside the body and also their biochemical reaction times are much shorter so can be excreted from the body before they start forming toxins. Another area where nanotechnology has been effectively used is drug capsulation and drug delivery. It has become possible now to selectively deliver drugs to selected cells in the body because of nano carriers like micelles.

In targeted drug delivery, the drug is released only within a targeted area of the body. (Yu et al., 2008) Therefore the therapeutic effect in diseased cells and/or tissues is maximized while side effects on normal tissues are minimized. Because of their unique pharmacokinetic properties, nanoparticles have a prolonged circulation time due to the fact that they are too large to be excreted by the kidneys in a short time, yet are small enough to avoid uptake by the reticulo endothelial system within the liver or spleen. Additionally, it has been shown that tumor vasculatures are not well developed, which allows nanoparticles with sizes smaller than 100 nm to pass through the endothelial cell layer and enter into the tumor mass. Delivery of therapeutic agents using nanoparticles offers a chance to improve the water solubility and bioavailability of hydrophobic drugs. Furthermore, because of their large surface area and diverse surface chemistry, nanoparticles can potentially be employed to combine drug therapies and facilitate

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 2. Nanotechnology regulates stem cell function. (a) Suitable spacing for integrin activation and integrin-mediated cell adhesion, cytoskeleton organization and cell spreading. (b) Modified nanoscaffolds regulate stem cell self-renewal, neural differentiation and osteogenic differentiation (Credit: Yang-Kao Wang).



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

the delivery of several distinct drugs simultaneously to a diseased tissue. (Wahajuddin, 2012) Targeting ligands can also be conjugated to the nanoparticle surface to facilitate selective and efficient delivery of drugs. Iron oxide-Dox (doxorubicin) nanoparticles can deliver higher levels of Dox into breast cancer cells and produce a stronger inhibitory effect on tumor cell growth. (Yu, 2008)

Nanomedicines, like antibody-drug conjugates (ADCs), gives a great potential of homing to the tumor where the exposure to the normal tissues is reduced. However, unlike the ADCs, which home to specific epitopes on cancer cells, nanomedicines exhibit a more dynamic form of tumor targeting, (Sengupta & Kulkarni, 2013) where increased circulation time, the enhanced permeability of tumor vasculature as compared to normal tissues, and the impaired tumor drainage contribute to increased intratumoral concentration of the nanoparticles as compared with free drugs. (Juliano, 2013) Nanomedicine will continue to evolve and will no doubt eventually find an important place in health care.

Regenerative Medicine

The application of nanotechnology in regenerative medicine has already begun to revolutionize several areas of stem cell research and will continue having great impacts on regenerative medicine.

Cells are single living units of organisms that receive the input signals from disease and injury and then return the output signals to their microenvironments. Conventional experimental studies for specific cellular responses are typically conducted on large cell populations, which inevitably produce data measured from an inhomogeneous distribution of cellular responses. Unless cellular responses and processes are isolated from inhomogeneous signals at the single cell level, it would be extremely difficult to elucidate the intricate cellular systems and to analyze the complex dynamic signaling transductions. Furthermore, conventional biomedical approaches reveal very little concerning genotypic aspects that transcend into cell phenotypes. Thus, to better understand and control the responses of cells towards external stimuli at the single cell or single molecule level, it is imperative to characterize the full range of cell behaviors (e.g., self-renewal, differentiation, migration and apoptosis). Figure 2 (Wu, 2013)

Recently, stem cells have gained much attention for the treatment of devastating injuries and damage caused by degenerative diseases, diabetes and aging. Stem cells self-renew for long periods of time and then further differentiate into specialized cells and tissues on stimulation by appropriate microenvironmental cues. They are typically categorized as embryonic stem cells (ESCs) or tissue-specific adult stem cells, depending on their origin and differentiation capability. ESCs, which originate from the inner-cell mass of the blastocyst-stage embryo are able to differentiate into all cell lineages found in the three primary germ layers of the embryo (e.g., endoderm, mesoderm and ectoderm) Although it has been shown that human ESCs (hESCs) can differentiate into many interesting cell types, such as cells of heart and brain.

The adhesion, growth and differentiation of stem cells are likely controlled by the surrounding microenvironment which contains both chemical and physical cues. Physical cues in the microenvironment, for example, nanotopography, were shown to play important roles in stem cell fate decisions. Thus, controlling stem cell behavior by nanoscale topography has become an important issue in stem cell biology. Nanotechnology has emerged as a new exciting field and research from this field has greatly advanced. Nanotechnology allows the manipulation of sophisticated surfaces/scaffolds which can mimic the cellular environment for regulating cellular behaviors. Thus, we summarize recent studies on nanotechnology with applications to stem cell biology, including the regulation of stem cell adhesion, growth, differentiation, tracking and imaging. Understanding the interactions of nanomaterials with stem cells

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

may provide the knowledge to apply to cell–scaffold combinations in tissue engineering and regenerative medicine (Solanki, 2008).

Chemistry

Here we are talking about the manufacture of new catalysts, coatings, computer components, highly selective sensors, lighter strong materials etc. The large surface to volume ratio will allow the development of new industrial catalysts. Nanomaterials can be used to make sensors that detect specific molecules. These detectors will find applications in detecting and monitoring pollutants in water. Nanotubes can be fabricated into strong and light materials, sometimes as composites with other non-nanomaterials. Such materials can be used in such diverse applications as aircraft construction (fuselage, wings etc.) and sports equipment. We are going to a few of these applications in detail and how nanotechnology has become an integral part of chemistry for creating such great materials.

George Whitesides says that the opportunities for chemistry to make important contributions to nanoscience abound. (Whitesides, 2005) Chemistry is unique in the sophistication of its ability to synthesize new forms of matter. The invention of new kinds of nanostructures will be crucial to the discovery of new phenomena. Materials science and chemistry are, over much of their shared border, indistinguishable. Chemistry has contributed (and will continue to contribute) to the invention and development of materials whose properties depend on nanoscale structure. Chemistry and chemical engineering will, ultimately, be important in producing these materials reproducibly, economically, and in quantity. (Whitesides, 2005) By understanding the molecular mechanisms of functional nanostructures in biology the light-harvesting apparatus of plants, ATPases, the ribosome, the structures that package DNA ultimately, the cell is an area where chemistry, with its singular understanding of molecular mechanism, can make unique contributions structures requires knowing what they are. Physical and analytical chemistry will help to build the tools that define these structures. Understanding the risks of nanostructures and nanomaterials will require cooperation across disciplines.

Engineering

The application of nanoscience and nanotechnology in engineering directly links academic research in nanoscience and nanotechnology to industries and daily life. As a result, numerous nanomaterials, nanodevices and nanosystems for various engineering purposes have been developed and used for human betterment. Here are a few examples of how an engineering aspect of nanotechnology has been used.

Tissue Engineering

Tissue engineering is an interdisciplinary field at the interface of engineering, materials science, medicine and biology. In typical tissue engineering approaches, cells are seeded onto a 3D biodegradable scaffold. As cells deposit their own matrix, the scaffold degrades, resulting in the formation of a biological tissue construct. In Tissue engineering, micro- and nanotechnologies can be used to fabricate biomimetic scaffolds with increased complexity and vascularization. (Goenka, Sant & Sant, 2014) Furthermore, these technologies can be used to control the cellular microenvironment (i.e., cell-cell, cell-matrix and cell-soluble factor interactions) in a reproducible manner and with high temporal and spatial resolution. Tissue and organ failure are serious and common medical conditions for which treatment options include

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

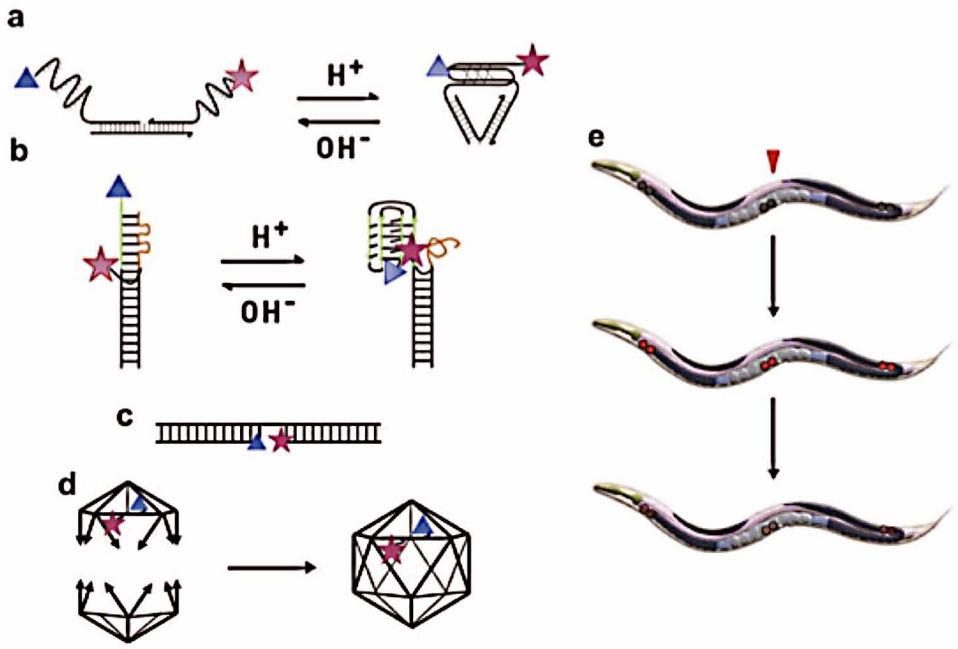
organ transplantation, surgical repair, artificial prostheses, and drug therapy. Transplantation is frequently hindered by the lack of tissue donors. To address this challenge, tissue engineering approaches are being developed to generate functional three-dimensional (3D) tissues. Efforts in this field may lead to the development of novel microscale platforms and nanomaterials that can help to solve today's problems of tissue engineering constructs.

DNA Nanotechnology

With all the above applications there is a whole new field emerged creating waves all around called DNA nanotechnology. Since DNA is a versatile scaffold for molecular sensing in living cells. Various cellular applications of DNA nanodevices have been demonstrated. DNA nanomachines are synthetic assemblies that switch between defined molecular conformations upon stimulation by external triggers. The invention of a DNA nanomachine called the *I-switch* has been a great advancement in the field of DNA nanotechnology and also in cell biology. The *I-switch* is a rationally designed, dynamic, DNA

Figure 3. Setting up an assay to assess the stability of DNA based nanostructures in vivo. Schematics illustrating (a) the structure and working of the three-stranded I-switch, $I_{A488/A647}$. (b) the principle of the re-engineered, two-stranded $I^{Tf}_{A488/A647}$. (c) a DNA duplex, $A_{TMR/A647}$, formed by three perfectly complementary oligonucleotides. (d) The formation of the DNA icosahedron ($IC_{FITC/A647}$) from two complementary half icosahedra. In each of these schematics, a blue triangle represents the donor fluorophore, while the acceptor is indicated by a red star. (e) The measurement of DNA architecture stability in vivo using coelomocytes of Caenorhabditis elegans, achieved by injecting the nematode with fluorescently labeled DNA nanostructures (red arrow) and then monitoring fluorescence decay in coelomocytes as a function of time.

Reproduced from: Surana, S., Bhatia, D., Krishnan, Y., (2013)Methods., 64, 94–100



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

based pH sensor, where a change in donor and FRET intensities is a read-out of the environmental acidity. *I-switch* is triggered by protons and functions as a pH sensor based on fluorescence resonance energy transfer (FRET) inside living cells. It is an efficient reporter of pH from pH 5.5 to 6.8, with a high dynamic range between pH 5.8 and 7.

The *I-switch* is a dynamic nanaodevice assembled in a one-pot synthesis and comprises only three DNA strands. It consists of two B-DNA duplexes connected to each other by a hinge, carrying terminal cytosine-rich single-strand overhangs. On acidification of the medium, the cytosine-rich over hangs protonate and form an I-motif thus causing a conformational change in the assembly. Figure 3. (Modi, Goswami, Gupta, Mayor, & Krishnan 2009)

CONCLUSION

In conclusion of the introductory chapter, we have introduced the basic applications of nanotechnology in the field of biomedicine, engineering and chemistry. In biomedicine considering that nanomaterials intrinsically enable cellular and molecular imaging with high sensitivity and high spatial resolution, it is not surprising that a growing number of imaging techniques based on nanoprobes are beginning to have an impact on stem cell-based therapies and research. Good examples of these include stem cell tracking using magnetic nanoparticles and QD-based imaging of stem cell interactions with other microenvironmental cues. In the filed of engineering though its vast we have picked up tissue engineering and DNA nanotechnology as the most emerging fields. For tissue engineering applications, improved biodegradable scaffolds are needed to provide cells with the proper signals to induce tissue formation.

In addition, other nanotechnology applications, such as surface engineering and drug delivery, have huge potential to further address challenges in the area of regenerative medicine. Collectively, advancements in nanotechnology enable the modulation of stem cell signaling pathways and improvement in their therapeutic applications. Finally, nanotechnology approaches for regenerative medicine essentially necessitate synergetic effort and interdisciplinary expertise from biology, chemistry and engineering. In particular, this approach would be beneficial to elucidate the complex cellular spatial–temporal dynamics and signaling pathways in more effective ways.

REFERENCES

Goenka, S., Sant, V., & Sant, S. (2014). Graphene-based nanomaterials for drug delivery and tissue engineering. *Journal of Controlled Release*, *173*, 75–88. doi:10.1016/j.jconrel.2013.10.017 PMID:24161530

Heidenreich, S., Büttner, H., & Ebert, F. (2003). Aerosole and ihre technische Bedeutung. *Chem. Ing. Techn.*, *75*(12), 1787–1809. doi:10.1002/cite.200306141

Iskandar, F. (2009). Nanoparticle processing for optical applications - A review. *Adv. Powder Techn.*, *20*(4), 283–292. doi:10.1016/j.apt.2009.07.001

Juliano, R. (2013). Nanomedicine: Is the wave cresting? *Nature Reviews. Drug Discovery*, *12*(3), 171–172. doi:10.1038/nrd3958 PMID:23449291

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Lenggoro, I. W., Okuyama, K., Fernández de la Mora, J., & Tohge, N. (2000). Preparation of ZnS nanoparticles by electrospray pyrolysis. *Journal of Aerosol Science*, *31*(1), 121–136. doi:10.1016/S0021-8502(99)00534-0

Liang, Z., Li, X., Xie, Y., & Liu, S. (2014). 'Smart' gold nanoshells for combined cancer chemotherapy and hyperthermia. *Biomedical Materials (Bristol, England)*, *9*(2), 025012. doi:10.1088/1748-6041/9/2/025012 PMID:24525482

Mijatovic, D., Eijkel, J. C. T., & Van den Berg, A. (2005). Technologies for nanofluidic systems: Top-down vs. bottom-up – a review. *Lab on a Chip*, *5*(5), 492–500. doi:10.1039/b416951d PMID:15856084

Modi, S., M. G, S., Goswami, D., Gupta, G. D., Mayor, S., & Krishnan, Y. (2009, May). A DNA nanomachine that maps spatial and temporal pH changes inside living cells. *Nature Nanotechnology*, *4*(5), 325–330. doi:10.1038/nnano.2009.83 PMID:19421220

Sengupta, S., & Kulkarni, A. (2013). Design principles for clinical efficacy of cancer nanomedicine: A look into the basics. *ACS Nano*, *7*(4), 2878–2882. doi:10.1021/nn4015399 PMID:23607425

Solanki, A., Kim, J., & Lee, K. (2008). Nanotechnology for regenerative medicine: Nanomaterials for stem cell imaging. *Nanomedicine (London, England)*, *3*(4), 567–578. PMID:18694318

Soni, S., Tyagi, H., Taylor, R. A., & Kumar, A. (2013). Role of optical coefficients and healthy tissue-sparing characteristics in gold nanorod-assisted thermal therapy. *International Journal of Hyperthermia*, *29*(1), 87–97. doi:10.3109/02656736.2012.753162 PMID:23311382

Wahajuddin, A., & Arora,. (2012). Superparamagnetic iron oxide nanoparticles: Magnetic nanoplatforms as drug carriers. *International Journal of Nanomedicine*, *7*, 3445–3471. doi:10.2147/IJN.S30320 PMID:22848170

Whitesides, G.M. (2005). Nanoscience, nanotechnology, and chemistry. *Small*, *1*(2), 172–179. doi:10.1002/smll.200400130 PMID:17193427

Wu, K.-C., Tseng, C.-L., Wu, C.-C., Kao, F.-C., Tu, Y.-K., C So, E., & Wang, Y.-K. (2013). Nanotechnology in the regulation of stem cell behavior. *Science and Technology of Advanced Materials*, *14*(5), 054401. doi:10.1088/1468-6996/14/5/054401

Yu, M. K., Jeong, Y. Y., Park, J., Park, S., Kim, J. W., & Min, J. J. et al. (2008). Drug-loaded superparamagnetic iron oxide nanoparticles for combined cancer imaging and therapy in vivo. *Angewandte Chemie (International ed. in English)*, *47*(29), 5362–5365. doi:10.1002/anie.200800857 PMID:18551493

KEY TERMS AND DEFINITIONS

Antibody-Drug Conjugates: They are monoclonal antibodies attached to biologically active drugs by chemical linkers with labile bonds designed for targeted therapy.

Dna Nanotechnology: DNA nanotechnology is the field in nanotechnology that uses the unique structure of DNA to create different structures.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Nanomaterials: The materials with at least one external dimension in the size range from approximately 1-100 nanometers.

Nanomedicine: One of the applications of nanotechnology in medicine is called Nanomedicine.

Regenerative Medicine: Regenerative medicine or cell therapy refers to treatments that are founded on the concept of producing new cells to replace malfunctioning or damaged cells as a vehicle to treat disease and injury.

Stem Cells: Stem cells are mother cells that have the potential to become any type of cell in the body.

Tissue Engineering: Tissue Engineering is the study of the growth of new connective tissues, or organs, from cells and a collagenous scaffold to produce a fully functional organ.

This work was previously published in the Handbook of Research on Diverse Applications of Nanotechnology in Biomedicine, Chemistry, and Engineering edited by Shivani Soni, Amandeep Salhotra, and Mrutyunjay Suar, pages 1-9, copyright year 2015 by Engineering Science Reference (an imprint of IGI Global).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 12
Nanotechnology Applications in Biomedical Engineering

Cajetan M. Akujuobi
Prairie View A&M University, USA

ABSTRACT

The 21st century has seen a massive explosion in the applications of nanotechnology. These applications cover all areas of Science, Technology, Engineering, and Mathematics (STEM). The advantage of nanotechnology comes from the fact that it has revolutionized the miniaturizations of many products that are useful to the well-being of society. A typical nanotechnology application example in biomedical engineering is its usage as drug eluting interfaces for implantable devices, such as vascular stents, orthopedic implants, and dental implants. The purpose of this chapter is to discuss the various applications of nanotechnology to biomedical engineering. Some of the future nanotechnology applications in biomedical engineering include healthcare/medical, consumer medical goods, environmental, and electronics. The impact of nanotechnology applications to biomedical engineering is in many ways enabling humans to survive different ailments that otherwise could have been very difficult to manage. The safety aspects in the applications of nanotechnology to biomedical engineering are also examined.

INTRODUCTION

The 21st century has seen diverse but very important applications of Nanotechnology. It is a field that has very long history and richness and its areas of applications are unlimited. It started with the first description of the concept of nanoscience in 1959 by Richard P. Feynman who was a physicist (Feynman, 1960). In 1974, Taniguchi (1974) came up with the term nanotechnology describing it as a field of engineering with precision having tolerances of micron or less. It was not until the mid1980s that Drexler (1986) with the publication of his book entitled "Engines of Creation" that the term nanotechnology became known to the public.

Nanotechnology applications cover all areas of science, technology, engineering and mathematics (STEM). It can be defined as that technology that has the ability to create using structures, devices and systems that have novel properties and functions because of their small and/or intermediate size. The

DOI: 10.4018/978-1-5225-0549-5.ch012

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

sizes (nanoscales) range from 1 nanometer (nm) to 100 (nm). Simply, nanotechnology means the world of "small things", such as atoms and molecules. It also concerns as well as nanoscience, very important materials called nanomaterials or nanotechnology materials.

The commercial applications as a result of the manipulations of these nanoscale materials include electronic, magnetic and optoelectronic, biomedical, pharmaceutical, cosmetic, energy, catalytic and materials applications. In this chapter, the focus will be on the nanotechnology applications to biomedical engineering. The specific application areas include sensors, "lab-on-chip" techniques, molecular imaging, drug delivery, biomedical electronics, biomedical nanorobots, etc. Nanotechnology applications in biomedical engineering presents many revolutionary opportunities in the fight against all kinds of cancer, cardiac and neurodegenerative disorders, infection and other diseases.

The applications of nanotechnology in biomedical engineering imply creation of materials and devices designed to interact with the body at sub-cellular scales with a high degree of specificity. This could be potentially translated into targeted cellular and tissue-specific clinical applications aimed at maximal therapeutic effects with very limited adverse-effects. The anticipated results include several direct products that are channeled to the improvements of diagnostics and therapeutics of the human body conditions.

The chapter will explore the different categories in the application of nanotechnology materials to the biomedical engineering field. The challenges facing scientists and engineers working in the field of nanotechnology will be discussed as they are quite enormous and extraordinarily complex in nature. The chapter is divided into eleven sections. Sections 1 and 2 cover the introduction and overview to nanotechnology applications while sections 3 and 4 cover the nanoscale materials and the nanotechnology particles respectively. Section 5 discusses the commercial applications of nanotechnology in biomedical engineering and in Section 6; we discuss the application of nanotechnology in biomedical engineering. In Sections 7 and 8; we discuss the nanotechnology safety issues and solutions and the future of nanotechnology applications in biomedical engineering respectively. Section 9 is the solutions and recommendations while Sections 10 and 11 are the future research directions and summary and conclusion and references respectively. Section 12 is the key terms and definitions.

OVERVIEW OF NANOTECHNOLOGY APPLICATIONS

In Nanoscience and Nanotechnologies (2004) and Knowles (2006), Nanometer, nanoscale and nanoscience are defined as follows: One thousand millionths of meter is called a *Nanometer (nm)*. An example is that of a human hair that is about 80,000 nm wide; a red blood cell is about 7,000 nm wide; and a water molecule is about 0.3 nm wide. In the case of *Nanoscale*, it is the size range from 100 nm down to 0.2 nm. The study of the phenomena and manipulation of materials at atomic, molecular and macromolecular scales where properties of the materials can be quite different from those of larger scales are called *Nanoscience*. Nanotechnology as well as nanoscience concern very important materials called *nanomaterials* or *nanotechnology materials*.

There are three different categories that will be explored in the application of nanotechnology materials in biomedical engineering. The first is the current applications that include coatings, clay, composites, tougher tools, and cosmetics. The second category is the short-term applications that include displays, fuel additives, paints, catalysts, batteries, and fuel cells. The third category is the long-term application of nanotechnology materials that include military battle suits, nanotube composites, water purification, machinable ceramics, medical implants, magnetic materials, and lubricants.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

In the biomedical engineering nanotechnology applications, diagnostic techniques, drugs, and prostheses and implants are three key areas of applications. The diagnostics nanotechnology sensors are used for doing things outside the body. In the analysis of blood and other samples, the "lab-on-chip" nanotechnology techniques are used. In the Research and Developments of new drugs for biomedical applications, the inclusion of nanotechnology techniques in biomedical analytical instruments becomes very important.

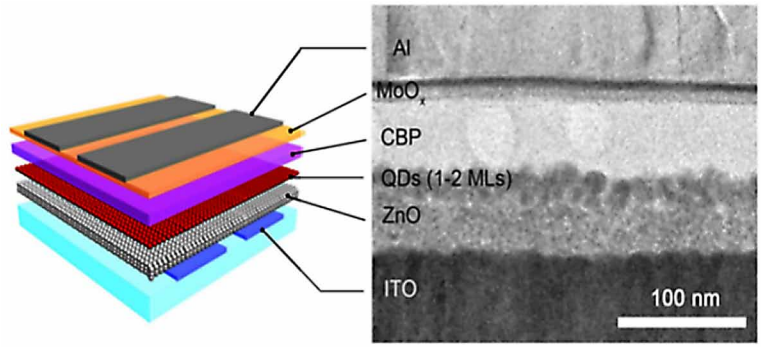
NANOSCALE MATERIALS

Nanoscale materials according to Walker (2006) can be defined as the set of substances that have at least one critical dimension less than 100 nanometers and possess unique optical, magnetic, or electrical properties. Typical examples of some of these nanoscale materials include metal oxides, fluorescent crystalline semiconductors (quantum dots), fullerenes, and carbon nanotubes.

In the nonmedical areas, other nanoscale materials existing commercially include nanoscale ceramics (nanoscale titanium dioxide or TiO2 and zinc oxide or ZnO) now used in certain cosmetics and sunscreens. Other manufactured nanoscale materials include fullerenes (C_{60}), single-walled carbon nanotubes (SWCNTs), multi-walled carbon nanotubes (MWCNTs), silver nanoparticles, iron nanoparticles, carbon black, titanium dioxide, aluminum oxide, cerium oxide, zinc oxide, silicon dioxide, polystyrene, dendrimers and nanoclays. Most of these materials do have direct applications in nanotechnology biomedical engineering of products such as in tissue engineering. Other biomedical engineering applications of nanoscale materials include sensors for medical and environmental monitoring, lab-on-a-chip diagnostic techniques, and longer lasting medical implants.

Nanoscale materials and engineering techniques have been used in such applications as boosting the performance of quantum dot light emitting diodes making the light at the end of the tunnel more efficient as shown in Figure 1. In the future, it may have direct applications to biomedical engineering devices due to its spectrally narrow, tunable emission, and ease of processing, that has made it an attractive material for LED technologies (Ambrosiano, 2006).

Figure 1. The quantum dot device structure shown with a transmission electron microscopy (TEM) image of a cross-section of a real device. Source: Los Alamos National Laboratory (www.lanl.gov), October 25, 2013.



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

The engineered nanomaterials have interesting applications to biomedical engineering. They tend to be man-designed and manufactured nanomaterials. Normally, they serve for a specific application purpose and can be found in different sizes/structures in different dimensions corresponding to wire-like, spheroidal and platelet-like characterized nanomaterials. Carbon and polymeric structures are examples of organic engineered types of nanometerials while metals, clays, and metalloid oxides including quantum dots are examples of the inorganic engineered types of nanomaterials.

NANOTECHNOLOGY PARTICLES

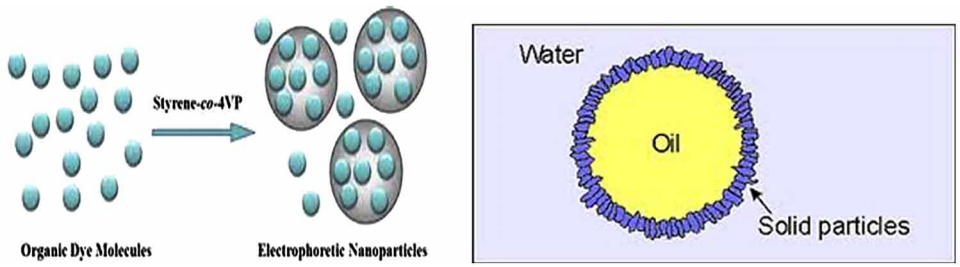
Nanotechnology particles are defined as small objects that behave as whole units with respect to their transports and properties. According to Nanoscience and Nanotechnology (2004), they can be also classified in terms of their different diameters. A range covering between 10,000 and 2,500 nanometers are called coarse particles while the range covering the sizes between 2,500 and 100 nanometers are called fine nanotechnology particles. The range of nanotechnology particles between 1 and 100 nanometers in size are called ultrafine particles, or nanoparticles.

While there are assorted types of nanotechnology particles, a good number of these nanotechnology particles are applied to biomedical engineering in so many ways. The list below demonstrates the importance of these nanotechnology particles to biomedical engineering. Figure 2 shows examples electrophoretic and oil solid nanotechnology particles that can be applied to biomedical engineering.

Quantum Dots: Also known as nanocrystals, quantum dots are nanosized semiconductors that, depending on their size, can emit light in all colors of the rainbow. These nanostructures confine conduction band electrons, valence band holes, or excitations in all three spacial directions. Examples of quantum dots are semiconductor nanocrystals and core-shell nanocrystals, where there is an interface between different semiconductor materials. They have been applied in biotechnology for cell labeling and imaging, particularly in cancer imaging studies.

Superparamagnetic Nanoparticles: Superparamagnetic molecules are those that are attracted to a magnetic field but do not retain residual magnetism after the field is removed. Nanoparticles of iron oxide with diameters in the 5 -100 nm range have been used for selective magnetic bioseparations. Typical techniques involve coating the particles with antibodies to cell-specific antigens, for separation from the surrounding matrix. Used in membrane transport studies, superparamagenetic iron oxide nanoparticles (SPION) are applied for drug delivery and gene transfection. Targeted delivery of drugs, bioactive mol-

Figure 2. Examples of Nanotechnology Particles. Source: http://www.google.com/images?



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

ecules or DNA vectors is dependent on the application of an external magnetic force that accelerates and directs their progress towards the target tissue. They are also useful as MRI contrast agents.

Nanorods: Typically 1-100 nm in length, nanorods are most often made from semiconducting materials and used in nanomedicine as imaging and contrast agents. Nanorods can be made by generating small cylinders of silicon, gold or inorganic phosphate, among other materials.

Fullerenes: The fullerenes, discovered in 1985 by researchers at Rice University, are a family of carbon allotropes named after Buckminster Fuller.

Liquid Crystals: Liquid crystal pharmaceuticals are composed of organic liquid crystal materials that mimic naturally-occurring biomolecules like proteins or lipids. They are considered a very safe method for drug delivery and can target specific areas of the body where tissues are inflamed, or where tumors are found.

Liposomes: Liposomes are lipid-based liquid crystals, used extensively in the pharmaceutical and cosmetic industries because of their capacity for breaking down inside cells once their delivery function has been met. Liposomes were the first engineered nanoparticles used for drug delivery but problems such as their propensity to fuse together in aqueous environments and release their payload, have led to replacement, or stabilization using newer alternative nanoparticles.

Nanoshells: Also referred to as core-shells, nanoshells are spherical cores of a particular compound surrounded by a shell or outer coating of another, which are a few nanometers thick.

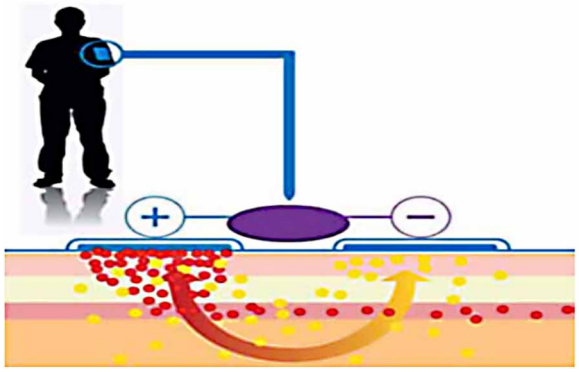
COMMERCIAL APPLICATIONS OF NANOTECHNOLOGY IN BIOMEDICAL ENGINEERING

The intent of any new technology is to commercialize it so that it will become a money making proposition. Nanotechnology is not an exception. On average, it takes anywhere from 3 to 10 years to get from research to product commercialization. That is why the corporations that get to the market faster than their competition generates more revenues that the others. The same case applies to biomedical engineering companies. In the nanotechnology applications to biomedical engineering, there are many commercialized products that are related to nanotechnology. Examples of commercialized applications of nanotechnology are in the manufacture of polymers based on molecular structure, and the design of computer chip layouts based on surface science. Taking advantage of the colloidal nanoparticles in bulk form as in cosmetics, protective coatings, drug delivery, and stain resistant clothing are the key to some of current commercial applications. Computers, medical devises, and sport equipment are some of the other consumer products in which nanotechnologies have been applied commercially. It is important to point out that in the industrialized countries such as the US, UK, Japan and Canada, consumer products and over 600 raw materials, intermediate components and industrial equipment items that are used by manufacturers contain nanomaterials. A great number of these nanotechnology produced commercialized products have direct applications to biomedical engineering.

According to Frederiksborgvej (2013), the global market share for nanotechnology applications is led by materials (31%), electronics (28%), medicine (17%), chemistry (9%), aviation (6%), and others (9%). The market for these technologies is expected to reach 2 billion US Dollars by year 2015, (Frederiksborgvej, 2013).This is interesting to note that most of these items like electronics, medicine, chemistry and nanomaterials have direct applications both commercially and otherwise to biomedical engineering. In figure 3, you will see an application of a nanotechnology based commercial electronics

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 3. Typical Iontophoresis Operation. Source: Smith, M.L. (Feb. 2012).



product device used for infusion drug therapy. The device is iontophoresis patch. According to Smith (2012), it uses electrical current to enable the infusion of a drug through the skin. The transdermal drug is ionized, dissolved in an aqueous solution and applied to an electrode in the patch. This specially formulated ionized compound can then be moved through the skin via direct current (DC) as shown in Figure 3. The nanotechnology technique made it possible for the miniaturization of the device so that it can be patched to the human.

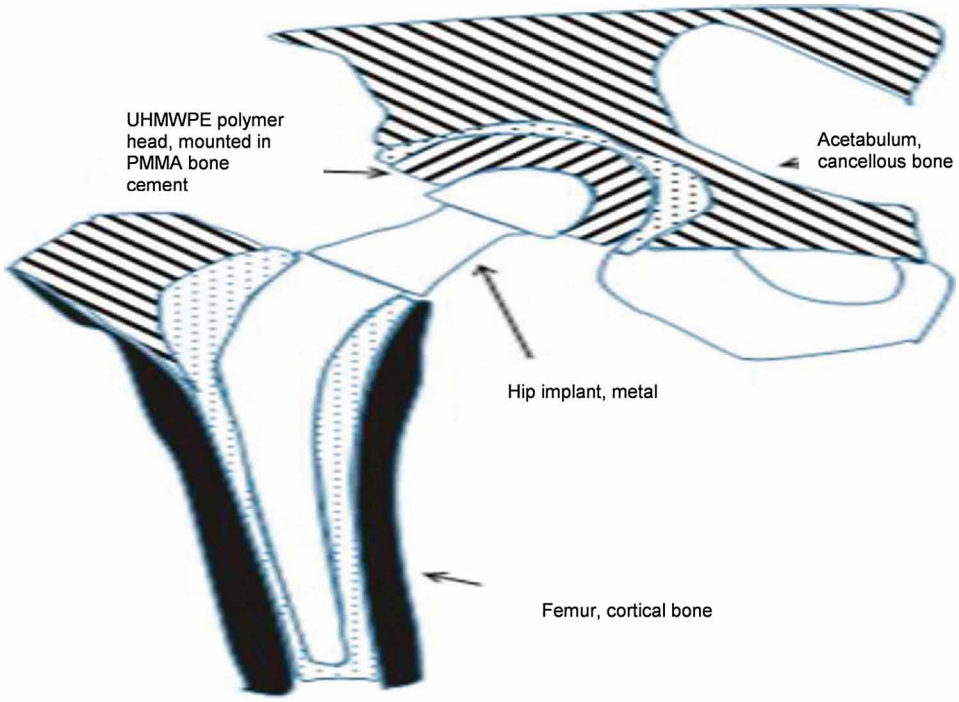
Iontophoresis has several advantages. One of the advantages is that the medicine can be locally administered at very high levels rather than being distributed throughout the body, which happens with syringe injections. This local administration can result in improved efficacy and reduced side effects. Advances in electronics technology, such as switched-mode power supply design, along with cost-effective high-performance microcontrollers have made the production of low-cost single-use dispensers for these drugs possible. Self-applied iontophoresis has already been used by many consumers to deliver medicines for many conditions including headaches, cold sores and wrinkles. These are very great nanotechnology-based commercially available products that have direct applications to biomedical engineering.

APPLICATIONS OF NANOTECHNOLOGY IN BIOMEDICAL ENGINEERING

Nanotechnology has been applied to biomedical engineering in so many ways in recent times. Interestingly, creating of controlled micro- and nanoarchitectures in an attempt to mimic the natural physical and biological environment that encourages tissue regeneration and growth is one such possible application. According to the nanotechnology researchers group (2013) with the Mechanical Engineering Department at Colorado State University; their hypothesis is that "the nanoarchitectures can promote cell differentiation and functionality". They also stated that "the ability to create model nanodimensional constructs that mimic physiological systems can aid in studying complex tissue interactions in terms of cell communication, response to matrix geometry, and effects of external chemical stimuli". A typical application of nanostructured materials in biomedical engineering is its usage as drug eluting interfaces for implantable devices, such as vascular stents, orthopedic implants, and dental implants. Controlling the size of nanoarchitecture, helps in manipulation of the release rates; thus releasing the drug at physiological levels.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 4. Schematic of a modern total hip replacement. Source: Blatchley (2013). p. 75.



The application of nanotechnology to biomedical engineering also extends to total hip replacement as shown in Figure 4. According to Blatchley (2013), the schematic shows a modern total hip replacement, mating a metal insert into the femur with an acetabular head made of ultra-high molecular weight (radiation cross-linked) polyethylene.

The tools necessary to understand and measure in the application to biomedical engineering is provided by nanotechnology means. These applications include manipulation of molecules within cells, biocompatible implants, smart controlled release of delivery of nutrients and biocompatible electronic devices as shown in Figure 5. This is an example of a nanotechnology in biomedical engineering-based electronics manufactured capsule injecting antibiotics to a cell that contains particularly targeted bacteria. This also extends to the delivery of drugs to cancer affected areas of the human body. In general, the nanotechnology applications to biomedical engineering include the various medical devices, tissue engineering, gene therapy, biotechnology, drug delivery and all aspects of biomedical pharmaceuticals as shown in Figure 6. The micro-camera serves as the eye of the device, the payload is carrying the medication to be delivered to the targeted area, the capacitor stores and provides the necessary power required to enable the device to travel to the required destination, while the swimming tail enables the device to float and navigate properly till it reaches the destination.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

NANOTECHNOLOGY SAFETY ISSUES AND SOLUTIONS IN THE BIOMEDICAL ENGINEERING

In the application of nanotechnology to biomedical engineering, even though there are many emerging great applications, there are many immediate safety concerns (Akujuobi, 2013). As an example, cell studies indicate there is genetic DNA damage that can be traced to nanotechnology particles in humans. The toxicity of nanotechnology particles aquatic wildlife cannot be ignored. Many species are at risk

Figure 5. Example of a nanotechnology in bioengineering-based electronics manufactured capsule injecting antibiotics to a cell that contains the targeted bacteria. Source: http://dm4community.com/?page_id=7

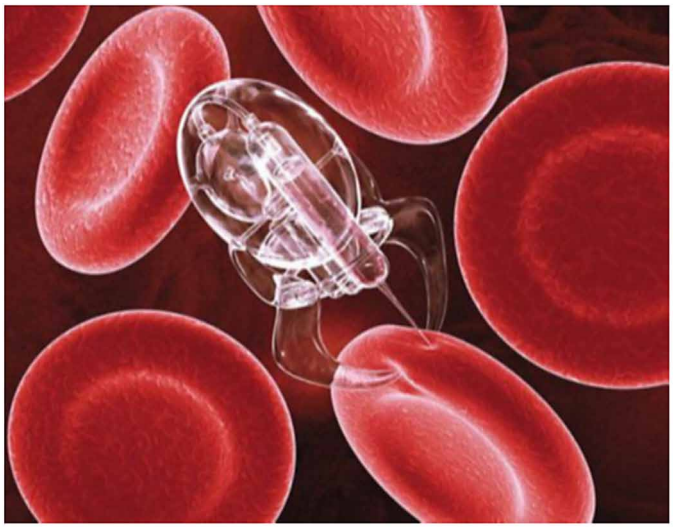
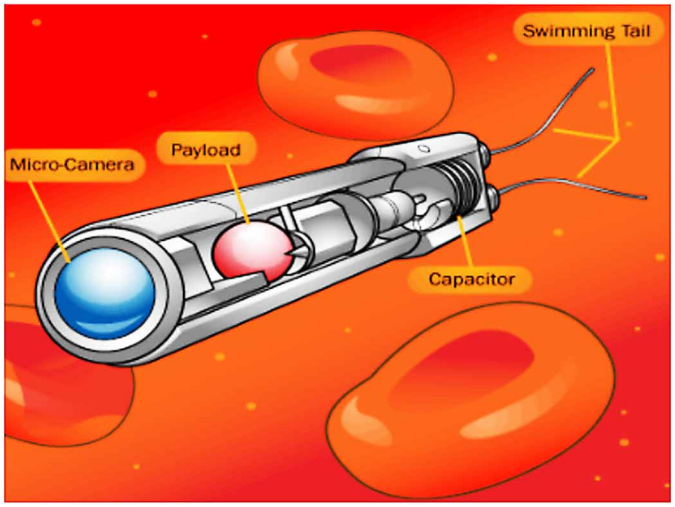


Figure 6. Example of the Concept of a nanorobot capable of delivering medicine straight to the virus location in the body via the blood stream. Source: http://electronics.howstuffworks.com/nanorobot.htm



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

and the global food chain highly threatened. It is imperative that as new applications in biomedical engineering continue to emerge and grow in the future, adequate attention therefore should be considered towards the safety solutions.

In the modern total hip replacement shown in Figure 3, wear particles from both the polymer and the metal have been found in surrounding tissue associated with inflammation and cement failure (Blatchley, 2013). This presents a major concern. Protective coatings incorporating nanomaterials tend to shift the distribution of these wear particles to smaller sizes. While there may be some solutions that can be explored, the total guarantee is not there yet to alleviate these concerns. However, all hopes are not lost as research continues to expand to find total solution.

The behaviors of nanoparticles are unlike lumps of the same materials, stronger, more toxic, and with radically different properties. The characteristic that makes them so useful also makes their safety so uncertain. As part of the safety solutions, there has to be a great consideration for immediate research into nanoparticle toxicities and its dissemination to the general public, especially, the communities that use nanotechnology particles such as the biomedical engineering communities. The effects of nanotechnology on the human health and the environment has resulted from the nanoparticles reaching waterways through wastewater treatment and disposal sites, thereby affecting the organisms that live in the water and the people who drink and cook with that contaminated water. This has an enormous effect on biomedical engineering and the biomedical products such as the different devices designed for human implants.

It is important to note that there are three types of nanoparticles that are of particular concern: carbon nanofibers; "buckyballs," or microscopic, football-shaped cages of carbon and nanosilver particles.

In tires carbon nanofibers are added, while in clothes, they are woven so as to produce different colors without using dyes. In the cases where they are shedded and inhaled, it could result to lung damage. In Li et al. (2009), it has been shown that a class of nanoparticles being widely developed in medicine called ployamidoamine dendrimers (PAMAMs) can cause lung damage by triggering a type of programmed cell death known as autophagic cell death. They have also shown that carbon-based "buckyballs" can be absorbed by organisms, thereby raising major concerns that the food chains can be contaminated by toxicities resulting to damages to humans.

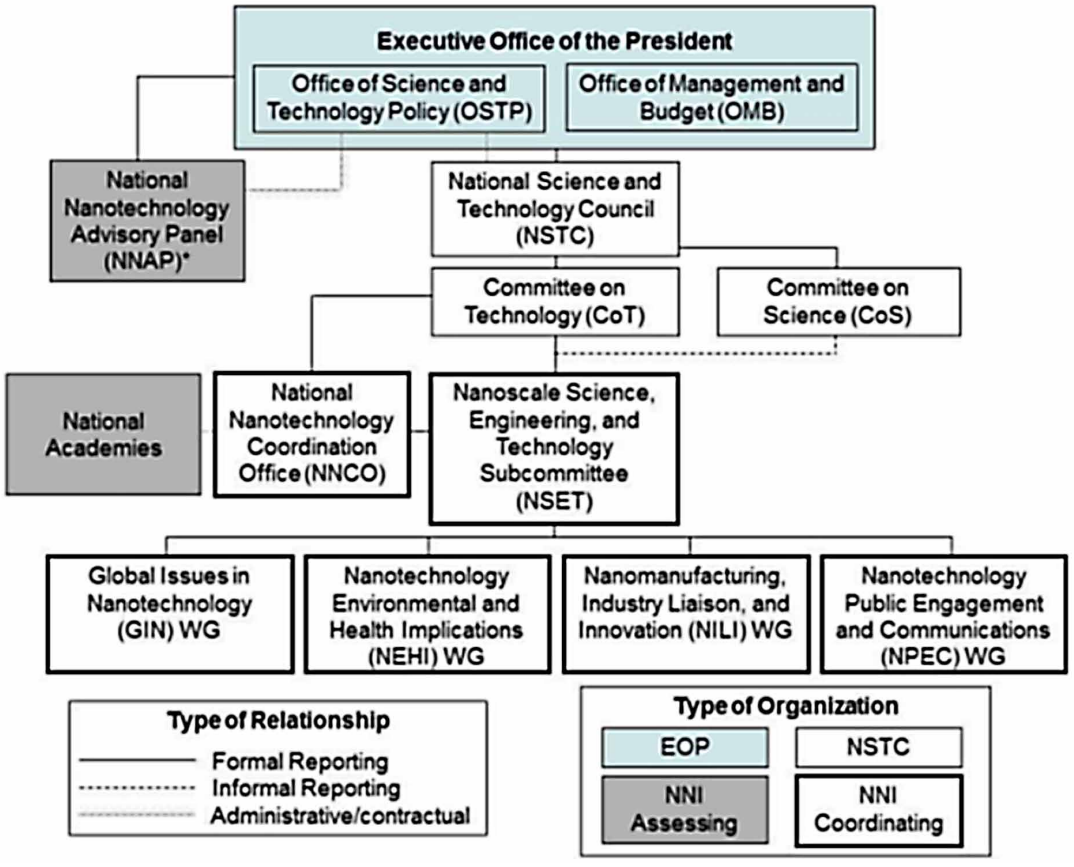
The other nanoparticle that may be of great concern to biomedical engineering applications is nanosilver. It is known to be highly toxic to aquatic life. While silver is safer for people than other toxics such as lead and chromium, for aquatic organisms, the story is quite different. Nanosilver is significantly more toxic than lumps of silver because the tiny particles' huge surface area increases their ability to interact with the environment. According to Geranio, Heuberger and Nowack (2009), nanosilver has been shown to break down and leach into water systems when, for example, sports garments incorporating silver nanoparticles for odor control are agitated in washing machines. In one of their studies, silver nanoparticles was used as antimicrobials in fabrics, of which out of the seven nanoparticle fabrics tested, four of them lost 20 percent to 35 percent of their silver in the first wash, and one brand lost half of its silver content in just two washings, all of which drained directly into the environment. It not surprising therefore, that many waterways are just recovering from high levels of silver introduced by the photography industry during the twentieth century. New silver nanoparticle products may result in highly toxic levels of silver being reintroduced into rivers and lakes through water treatment facilities.

The use of nanotechnology in consumer products continues to grow rapidly. More than one thousand nanotechnology-enabled products have been made available to consumers around the world. Examples of such products include biomedical engineered products and other non-biomedical engineered products that may have tiny particles in everything from conventional products like non-stick cookware and

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

lighter, stronger tennis racquets, to more unique items such as wearable sensors that monitor posture. In 2009, more than 600 different known products were manufactured using nanotechnology commercially (Hutchinson, 2009). As of March 10, 2011, that number had grown to about 1,300 products; and it is anticipated that by year 2020, it should be about 3,400 (The Project on Emerging Nanotechnologies, 2011). If not properly controlled, all these new emerging nanotechnology products applied to biomedical engineering and other areas will provide significant oversight challenges for agencies like the Food and Drug Administration and Consumer Product Safety Commission, which often lack the necessary mechanisms to identify unsafe nanotechnology products before they enter the marketplace. However, organizations such as the ones shown in n Figure 7 can work towards finding the safety solutions just as they help with the nanotechnology initiatives of the nation.

Figure 7. Organizations with a Role in the National Nanotechnology Initiative and Their Relationships. Source: Reproduced from The National Nanotechnology Strategic Plan, Nanoscale, Science, Engineering, and Technology Subcommittee, National Science and Technology Council, The White House, February 2011.



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

THE FUTURE OF NANOTECHNOLOGY APPLICATIONS IN BIOMEDICAL ENGINEERING

Most of the future applications of nanotechnology applications in biomedical engineering can be grouped under the following areas, healthcare/medical, consumer medical goods, environmental and electronics, power/energy and engineering. Biomedical engineering nanostructures that include coatings and thin films are possible areas that have great future. Renn and Roco (2006) have shown that there are about four generations of nanotechnology that projects the past, present and future. The first generation that took place approximately up to the year 2000 had to do with the passive nanostructures. It covered the dispersed and contact nanostructures and the products that incorporate the nanostructures. Within the year 2005, the second generation active nanostructures came into existence. These were the bio-active and health effects of nanostructures. The third generation, the various systems of nanosystems started around 2010 given rise to guided assembling, 3D networking and new hierarchical architectures. Finally, the fourth generation deals with the molecular nanosystems that comprises the molecular devices 'by design', atomic design and emerging functions. These fourth generation nanotechnology covers the periods up to 2015-2020.

The future of nanotechnology in the biomedical engineering world will see nanotechnology leading the way in the fabrication of biomedical devices that will be cleaner, stronger, lighter and more precise. It will also lead the way to seeing future biomedical engineering devices that can achieve Gigabit networks. These types of networks will be needed in handling vast amounts of voice, image, and video data being created and will be instrumental in the biomedical field. Should wireless devices be needed in the future of nanotechnology applications to biomedical engineering, components based on the microelectromechanical systems (MEMS) and nanoelectromechanical systems (NEMS) do hold significant promise for future developments of wireless biomedical devices.

SOLUTIONS AND RECOMMENDATIONS

The applications of nanotechnology to biomedical engineering will continue to expand in leaps and bounds in years to come. The applications are not only limited to biomedical engineering but to many other areas of human endeavor. However, listed below are some specific areas where the research trends seem to drive the applications as far as biomedical engineering is concerned. This author is recommending that the research and applications should continue to grow and be commercialized with all the safety concerns taking into consideration.

- The drug delivery issues have already been improved in the laboratory environments through nanocapsules and nanotubes. This application should be encouraged to expand and efficiently applied as it would increase the treatment and recovery rates of cancer patients without damaging healthy cells.
- The facilitation of using synthetic peptides (Antimicrobial peptides) that kills bacteria while destroying less blood cells is an area that the nanotechnology applications to biomedical engineering should continue to expand and grow to commercialization.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

- The applications of nanotechnology biomedical engineering to the idea of identifying cancerous molecules and in addition to nucleic acids belonging to specific strains of cancer such as different types of lung and breast cancers should be encouraged to continue.
- The exploration of the use of gold as a catalyst to remove a step in the production of pharmaceutical drugs should be encouraged to continue both in the laboratory experiments and in the commercialization of the products. This tends to lead to higher yield of products (Gu, Neo and Zhang, 2011).
- The application of nanotechnology to biomedical engineering activities such as the diagnostics tailored to an individual by identifying particular nucleic acids or identified genes should be encouraged to expand and continue.
- In the application to HIV/AIDS, using nanotechnology to transport agents targeted to the desired site thereby evading the body's immune system should be allowed to expand.
- The monitoring of diabetic conditions using nanotechnology applications to biomedical engineering, acts as an artificial means of regulating and maintaining the body's own hormonal balance in the war against diabetes which is very important.
- In the contraception world, using the nanotechnology to transport agents targeted to the desired site with minimal side effects should be encouraged to expand and not languish in laboratory trials.

FUTURE RESEARCH DIRECTIONS

In the nanotechnology world, there great emerging trends and innovations that are leading to the way to the obvious future directions of research considerations. Specifically in the area of biomedical engineering, tissue engineering research has a great potential future. For those patients who may need new bones, new teeth, new limb, how can you help them solve their problems efficiently and at an affordable cost? These are research problems that those researchers in the business of coming up with research solutions in better prostheses and implants should find improved and efficient solutions. Drug delivery mechanisms are another area of future research. Significant progress has been made, but the question still remains as to the extent in which these nanotechnology-enabled systems can move around most parts of the body more efficiently and more cost effective. Finally, safety is a major area of concern. Are there safety measures in place that can effectively give comfort to the nanotechnology research community and to the users of nanotechnology products? More research may be needed in this direction.

SUMMARY AND CONCLUSION

Even though we are focusing on the applications of nanotechnology in the biomedical engineering world, this field of smaller devices has an enormous potential and the future looks very bright. The question therefore is, how far can this go in this 21st century and how can we make sure that the safety effects are thoroughly studied, understood and implemented alongside the inevitable applications to the biomedical engineering and other areas. In the commercialization business, there seems to be a terrible gab between research and commercialization and this has to be addressed by both the venture capitalists and the government agencies. The impact of nanotechnology applications to biomedical engineering is,

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

in many ways enabling humans to survive different ailments that otherwise could have been very difficult to manage. As nanotechnology makes biomedical devices increasingly smaller and more efficient, collecting, storing, sharing and processing large amounts of information easily in the drug delivery and diagnostics areas will become easier and cheaper.

REFERENCES

Akujuobi, C. M. (2013). Nanotechnology safety in the electronics and telecommunication industries. In Nanotechnology safety. Elsevier. doi:10.1016/B978-0-444-59438-9.00011-4

Ambrosiano, N. (2013). *Nanoscale engineering boosts performance of quantum dot light emitting diodes - Making the light at the end of the tunnel more efficient*. Los Alamos National Laboratory. Retrieved from www.lanl.gov

Blatchley, C. (2013). Nanotechnology Safety in the Biomedical Industry. In R. Asmatulu (Ed.), *Nanotechnology safety* (pp. 73–83). Amsterdam: Elsevier. doi:10.1016/B978-0-444-59438-9.00006-0

Drexler, E. (1986). *Engines of creation: The coming era of nanotechnology*. Retrieved from http://www.nanotec.org.uk/finalReport.htm

Feynman, R. P. (1960). *There is plenty of room at the bottom. California Institute of Technology Journal of Engineering Science*, 4, 23-36.

Frederiksborgvej. (2013). *Nanotechnology*. Riso National Laboratory.

Geranio, L., Heuberger, M., & Nowack, B. (2009). The behavior of silver nanotextiles during washing. *Environmental Science and Technology, 43*(21), 8113–8118.

Gu, L., Neo, B. S., & Zhang, Y. (2011). Gold-catalyzed direct amination of arenes with Azodicarboxylates. *Organic Letters*, *13*(7), 1872–1874. doi:10.1021/ol200373q PMID:21366340

Hutchinson, T. (2009). *Safety concerns over nanotechnology, The 7:30 Report*. Retrieved from www.abc.net.au/7.30/content/2009/s2592535.htm

Knowles, E. E., III. (2006). *Nontechnology: Evolving Occupational Safety, Health and Environmental Issues*. Retrieved from http://www.redorbit.com/news/science/436575/

Li, C., Liu, H., Sun, Y., Wang, H., Guo, F., & Rao, S. et al. (2009). PAMAM Nanoparticles Promote Acute Lung Injury by Inducing Autophagic Cell Death through the Akt-TSC2-mTOR Signaling Pathway. *Journal of Molecular Cell Biology*, *1*(1), 37–45. doi:10.1093/jmcb/mjp002 PMID:19516051

Nanotechnology Researchers Group. (2013). *Nanotechnology for Biomedical Applications*. Retrieved from http://www.engr.colostate.edu/me/pages/research/health/Nanotechnology.html

Project on Emerging Nanotechnologies. (2011). Retrieved from www.nanotechproject.org/news/archive/9231/

Renn, O., & Roco, M. C. (2006). Nanotechnology and the Need for Risk Governance. *Journal of Nanoparticle Research*, *8*(2), 153–191. doi:10.1007/s11051-006-9092-7

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Smith, M. L. (2012). *Designing Cutting-Edge Wearable Medical Devices*. Retrieved from http://electronicsforu.com/electronicsforu/circuitarchives/view_article.asp?sno=773&title%20=%20Designing+Cutting%2DEdge+Wearable+Medical+Devices&id=12036&article_type=8&b_type=new

N. Taniguchi (Ed.). (1974). *Proceedings of international conference on production engineering* (vol. 2). Tokyo, Japan: Society of Precision Engineering.

Walker, N. (2006). *NTP Nanotechnology Safety Initiative, NIEHS/NIH*. Retrieved from http://ntp-server.niehs.nih.gov/ntp/Factsheets/NanoColor06SRCH.pdf

KEY TERMS AND DEFINITIONS

Biomedical Engineering: It is the study and application of engineering principles and design to biological and medical sciences.

Nanometer (nm): One thousand millionths of A meter is called a nanometer.

Nanoscale: It is the size range from 100 nm down to 0.2 nm.

Nanoscience: The study of the phenomena and manipulation of materials at atomic, molecular and macromolecular scales where properties of the materials can be quite different from those of larger scales.

Nanotechnology Particles: They are defined as small objects that behave as whole units with respect to their transports and properties.

Nanotechnology: The ability to engineer functional systems at the molecular scale.

Superparamagnetic Nanoparticles: These are nanoparticles that are attracted to a magnetic field but once the magnetic field is removed, the residual magnetism is not retained.

This work was previously published in the Handbook of Research on Diverse Applications of Nanotechnology in Biomedicine, Chemistry, and Engineering edited by Shivani Soni, Amandeep Salhotra, and Mrutyunjay Suar, pages 50-63, copyright year 2015 by Engineering Science Reference (an imprint of IGI Global).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 13
Nanotechnology for Omics-Based Ocular Drug Delivery

Anjali Hirani
University of South Florida, USA & Virginia Tech-Wake Forest University, USA

Yong Woo Lee
Virginia Tech-Wake Forest University, USA

Yashwant Pathak
University of South Florida, USA

Aditya Grover
University of South Florida, USA

Vijaykumar Sutariya
University of South Florida, USA

ABSTRACT

Millions of people suffer from ocular diseases that impair vision and can lead to blindness. Advances in genomics and proteomics have revealed a number of different molecular markers specific for different ocular diseases, thereby optimizing the processes of drug development and discovery. Nanotechnology can increase the throughput of data obtained in omics-based studies and allows for more sensitive diagnostic techniques as more efficient drug delivery systems. Biocompatible and biodegradable nanomaterials developed through omics-based research are able to target reported molecular markers for different ocular diseases and offer novel alternatives to conventional drug therapy. In this chapter, the authors review the pathophysiology, current genomic and proteomic information, and current nanomaterial-based therapies of four ocular diseases: glaucoma, uveal melanoma, age-related macular degeneration, and diabetic retinopathy. Omics-based research can be used to elucidate specific genes and proteins and develop novel nanomedicine formulations to prevent, halt, or cure ocular diseases at the transcriptional or translational level.

OCULAR DISEASE

Approximately 140 million Americans over the age of 40 suffer from a variety of ocular diseases that impair vision and may lead to blindness (NEI, 2012). The prevalence of ocular disorders will continue to increase with the worldwide aging population. Although current treatments do exist, there is a need

DOI: 10.4018/978-1-5225-0549-5.ch013

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

for better diagnostics and more efficient therapies that can arrest the progression and/or even reverse damage of ocular diseases. Currently, more information is needed to understand the pathogenesis of ocular disease as well as determining new drug targets to enhance ocular drug discovery or repositioning current drugs for more efficacious therapy.

Limitations in Treatment for Ocular Diseases

Currently, therapies exist to delay progression of some ocular diseases; however, a better understanding of pathogenic processes is needed to find more effective treatments. Some of the common ocular diseases presented later in this chapter possess a complex interplay of genetic factors that are challenging to treat. Discovery of genes responsible for ocular disorders can aid in the development of new therapeutic agents. Clinically, we are merely treating symptoms and not targeting the actual disease mechanisms. More research needs to be completed to elucidate these mechanisms.

A number of barriers for ocular drug delivery exist, such as nasolacrimal drainage and the blood-aqueous and blood-retinal barriers. This restricts administration of potential therapeutics. Drug can be delivered by a variety of routes including topical ocular, periocular injection, intravitreal injection, and systemic administration. The topical route is a convenient method of drug delivery to the anterior segment; however, a model of transient diffusion has shown that less than 5% of a lipophilic drug and 0.5% of a hydrophilic drug penetrate the cornea (W. Zhang, Prausnitz, & Edwards, 2004), and the remainder is cleared through nasolacrimal drainage and systemic absorption(Gambhire, Bhalerao, & Singh, 2013). The amount of available drug that permeates across the sclera is reduced with cationic and lipophilic solutes and the RPE has tight intercellular junctions for hydrophilic molecules (Urtti, 2006). Additionally, the lymphatic system, blood vessels and active transporters all work to clear drugs administered through transscleral routes. Systemic administration of drugs requires high doses that are potentially toxic to obtain a therapeutic concentration across the blood ocular barriers (Geroski & Edelhauser, 2000; Sigurdsson, Konraethsdottir, Loftsson, & Stefansson, 2007). Intravitreal injections circumvent physiological barriers and maintain therapeutic doses without damage to bystander tissues; however, frequent injections can lead to complications like retinal detachment, increase in ocular pressure, and hemorrhage (Peyman, Lad, & Moshfeghi, 2009). Given the presence of these physiological barriers, the development of therapies utilizing nanotechnology that efficiently deliver drugs and extend drug release to the eye would be beneficial to the progression of ocular disease treatment. Due to the lengthy pipeline in gaining FDA approval, newly repositioned drugs can be utilized to expand current disease therapy.

Omics-Based Nanotechnology

The goal of omics-based nanotechnology is to use nanoscale technology to enhance early detection, gain an understanding of pathophysiology, as well as find better treatments for eye disease. Nanogenomics refers to a new approach for medical diagnostics and therapy (Nicolini, 2006). Nanoproteomics allows us to evaluate expression of ocular proteins, identify novel therapeutic targets study the pharmacological effects of therapeutics (Steely & Clark, 2000). These new techniques can improve the current understanding of ocular diseases and aid in the discovery of therapies targeting genes responsible for ocular disease.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

APPLICATION TO OCULAR DISEASES

Glaucoma

Pathophysiology

Glaucoma causes blindness in about 7 million people every year, only 10% of the people affected by glaucoma worldwide (Quigley, 1996). A number of risk factors have been identified for glaucoma, including hypertension, family history, and age, among others (Abbot F Clark & Thomas Yorio, 2003). Glaucoma is caused by progressive damage to the trabecular meshwork, optic nerve head, and retinal ganglion cells; however, the exact mechanism behind the nerve damage is yet unknown (Abbot F Clark & Thomas Yorio, 2003; Frank S Ong et al., 2013). Glaucoma can usually be diagnosed as open- or closed-angle and primary or secondary based on the degree of ocular hypertension (Abbot F Clark & Thomas Yorio, 2003). Elevated intraocular pressure (IOP) is characteristic of primary open-angle glaucoma (POAG), the most common type of glaucoma, and is the most common target of treatment.

Current Genomic and Proteomic Information

A number of genetic markers that play a role in multiple different types of glaucomas have been identified.

MYOC was one of the first genes identified for juvenile and adult-onset POAG (Stone et al., 1997). Its associated protein, myocilin, has been associated with the trabecular meshwork and may be associated with increased outflow resistance, leading to an increase in IOP pressure (Fautsch, Bahler, Jewison, & Johnson, 2000; WuDunn, 2002; Z. Zhou & Vollrath, 1999).

Reiger syndrome is a genetic disorder which manifests itself in craniofacial and umbilical abnormalities (Amendt, Semina, & Alward, 2000). The varied phenotypes associated with Reiger syndrome are associated with different mutations in the *PITX2* gene and the PITX2 transcription factor for which it codes. More than half of the individuals diagnosed with Reiger syndrome develop glaucoma, making the mutant *PITX2* gene a risk factor and molecular marker for glaucoma (WuDunn, 2002).

Primary congenital glaucoma has been traced to mutations in *CYP1B1* gene. *CYP1B1* codes for the CYP1B1 protein from the cytochrome P450 family, but its mechanism in ocular pathology is not yet fully understood (Stoilov, Jansson, Sarfarazi, & Schenkman, 2001). Studies of various ethnic populations affected with congenital glaucoma revealed multiple mutations of the *CYP1B1* gene responsible for the disease in Western- and Middle Eastern as well as Japanese populations (Michels-Rautenstrauss et al., 2001).

Six genes (*GLC1A*-*GLC1F*) have been identified through pedigree analysis of families affected by hereditary POAG. *PCOLCE2* has also been identified as a possible indicator of glaucoma because of its association with the trabecular meshwork and the *GLC1C* locus, but no mutations in *PCOLCE2* were identified in patients with *GLC1C*-induced POAG (WuDunn, 2002). Population studies in Estonian POAG patients also showed a significantly higher association between individuals positive for the glutathione S-transferase (GSTM1) protein with POAG as compared to a control population. Smoking increased this risk (Juronen et al., 2000).

Reductions in the outflow of aqueous humor from the anterior chamber of the eye have been implicated in glaucoma pathology due to its association to IOP. Imbalances in the expression levels of matrix metalloproteinases (MMPs) and tissue inhibitors of MMPs (TIMPs), which remodel the extracellular

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

matrix (ECM) of the trabecular meshwork, may account for the differences in aqueous humor outflow in patients with glaucoma (Wong et al., 2002). Disruption of trabecular meshwork ECM was also revealed as a result of increased TGF-beta levels in glaucoma patients (Zhao, Ramsey, Stephan, & Russell, 2004). Prostaglandins increase the activity of MMPs in ciliary smooth muscle cells, which effects the outflow of aqueous humor, and can lead to a decrease in the high IOP pressure characteristic of POAG (Weinreb, Kashiwagi, Kashiwagi, Tsukahara, & Lindsey, 1997). In addition, cochlin protein has been identified as an elevator of IOP pressure in glaucoma patients by disrupting the trabecular meshwork (Bhattacharya et al., 2005).

Current Therapies

A number of current nanomedicine therapies have been investigated in the treatment of glaucoma physiology. The benefits that nanomedicines offer are their biodegradability, sustained release of drugs, and targeting of necessary tissues and molecular pathways.

Chu et al. investigated the activity of 7-hydroxy-2-dipropylaminotetralin (7-OH-DPAT)-loaded calcium phosphate nanoparticles (CAP) in the reduction of IOP and aqueous flow rate in glaucomatic rabbits. 7-OH-DPAT in CAP showed a hypotensive, therefore therapeutic, effect in the glaucomatic model; however, raclopride, a dopamine D2/D3 receptor agonist, was shown to reduce the effect of 7-OH-DPAT in CAP. The data suggests that dopamine D2/D3 receptors may play a role in modulating IOP and that CAP may potentially be a therapeutic agent for glaucoma. In addition, further investigations of the role of the dopamine D2/D3 receptor in glaucoma may elucidate further directed therapies for the modulation of glaucomatic IOP (Chu, He, & Potter, 2002).

Wadhwa et al. revealed the IOP-reducing effects of hyaluronic acid (HA)-conjugated chitosan (CS) nanoparticles (CS-HP-NPs) loaded with dorzolamide hydrochloride and timolol maleate, 2 drugs used in glaucoma treatment. CS-NPs have been used as nanomedicine carriers due to their biodegrability and drug availability, but conjugation with HA synergistically increases the mucoadhesion of CS-NPs. *In vivo* studies in rabbits revealed a significant IOP reduction in rabbits treated with CS-HA-NPs as opposed to the free drug solution, suggesting its potential use against glaucoma (Wadhwa, Paliwal, Paliwal, & Vyas, 2010).

Jiang et al. reported the neuroprotective role of glial cell-line derived neurotrophic factor (GDNF) loaded into poly DL-lactide-co-glycolide (PLGA) microspheres in elevated IOP-induced rats. Intravitreal injections of the GDNF-PLGA microspheres were significantly better in the neuroprotection of retinal ganglion cells (RGCs) in chronically elevated IOP through the analysis of glial fibrillary acidic protein (GFAP) production compared to the blank PLGA microspheres and GDNF alone (Jiang et al., 2007). It is reasonable to believe that the nanomedicine administration of various other genes and RNA inhibitors of proteins may be able to provide further neuroprotection to RGCs after injury or surgery (Zarbin, Montemagno, Leary, & Ritch, 2013). Indeed, as a potential post-operative therapy to protect against scarring after glaucoma surgery, Dos Santos et al. investigated the role of nanosized complexes of antisense TGF-beta2 phosphorothiorate oligonucleotides (PS-ODN) with polyethylenimine (PEI) encapsulated in PLGA microspheres. Subconjunctival injections in rabbits significantly improved bleb survival and increased the intracellular penetration of PS-ODN, revealing the microspheres' potential therapeutic role (Gomes dos Santos et al., 2006).

The incorporation of drugs, genes, or oligonucleotides into biodegradable and biocompatible nanomedicine vectors may provide novel therapeutic potentials in the prevention and treatment of glaucoma.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

The size range of nanomedicines make these vectors compatible for endocytosis into effected cells and also offer sustained release of the compounds associated with the vectors. These compounds may be able to directly influence the genetic and molecular pathways that influence the glaucoma phenotype.

Uveal Melanoma

Pathophysiology

Uveal melanoma (UM) is one of the most common ocular cancers in adults, with 7 cases per million each year and over 20 cases per million in those over the age of 70 (Ramasamy et al., 2014; Singh & Topham, 2003). Over ninety percent of UM cases arise in the choroid and have the worst prognosis, followed by the ciliary body and iris, with the best prognosis (Ramasamy et al., 2014). Over half of UM cases result in metastasis, 40% of which result in death even after primary tumor treatment (Bedikian, 2006). Close to 90% of UM metastatic cases result in metastasis to the liver; the skin, bones, and lungs are other common sites of metastasis following metastasis to the liver (Lorigan, Wallace, & Mavligit, 1991; Singh & Borden, 2005).

Current Genomic and Proteomic Information

Heat shock protein 27 (HSP-27) is a cytoplasmic protein involved in cell migration, cytoskeletal structure, cell survival, and tumor progression (Kostenko & Moens, 2009). It plays different roles in a number of different cancers; HSP-27 is over expressed in gastric, prostate, and node-negative breast carcinoma and indicates a poor prognosis in each, but over-expression of HSP-27 indicates good prognosis in non-small-cell lung- and ovarian carcinomas (Ramasamy et al., 2014). Proteomic analysis of primary UM tissue revealed a down regulation of HSP-27 in monosomy 3 tumors, with a significantly lower expression in monosomy 3 tumors as opposed to disomy 3 tumors (Coupland et al., 2010). In addition, Jmor et al. showed that a significantly reduced expression of HSP-27 in UM tissue correlated with a predicted survival of less than 8 years (Jmor, Kalirai, Taktak, Damato, & Coupland, 2012). Investigations in a human cuteanous melanoma cell line revealed that the over expression of HSP-27 inhibited cell proliferation and reduced cell invasiveness, which suggests that the under expression of HSP-27 might induce greater cell migration in UM (Aldrian et al., 2002).

Two-dimensional difference gel electrophoresis (2D DIGE) followed up by immunohistochemical studies in primary UM tumor samples also identified the over expression of fatty acid-binding protein heart type (FABP3) and triosephosphate isomerase (TPI1). In addition, siRNA knockdowns of these 2 proteins in a primary UM cell line revealed significantly reduced levels of cell invasion and migration (Linge et al., 2012). FABPs are thought to play a number of metabolic roles in cells, including growth, differentiation, and apoptosis (Lichtenfels et al., 2009). Investigation of the over expression of FABP in small cell lung cancer implicate its role in mitosis and cell growth (L. Zhang, Cilley, & Chinoy, 2000). FABP has also been linked with poor prognosis and tumor aggressiveness in human gastric carcinoma (Hashimoto et al., 2004). TPI1 plays a role in the cell's glycolysis and gluconeogenesis pathways, high rates of which are required for tumor cells (Albery & Knowles, 1976; Bui & Thompson, 2006). TPI1 has been reported to be associated with the aggressiveness of breast cancer, and its under expression was found to induce apoptosis in the HeLa cell line (Lee et al., 2010; Selicharova et al., 2008). However, its over expression in lung cancer tissue, cell lines, and plasma, as well as in prostate cancer implicates its

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

role in the progression of disease and as biomarkers (Chen et al., 2002; Kim, Koo, Kim, Sohn, & Park, 2008; Qian et al., 2010). The identification of these genetic markers in UM and their varied roles in a number of different types of cancers suggests that their over expression in primary UM tumors may contribute to the oncogeneity of the tissue. Further investigations of the roles of FABP3 and TPI1 in UM may reveal novel pathways for therapeutic intervention as well as screening techniques in preventing the progression of UM.

Proteomic analyses of primary tumors and cell lines have revealed the above-mentioned proteins as potential biomarkers for UM; however, further such studies are required to elucidate more proteins for the identification of UM. There is a lack of genomic and proteomic analysis of UM tissues in available literature, but such studies would greatly increase the understanding of the molecular mechanisms behind the induction and progression of the disease. The proteins and biomarkers identified by future studies would provide researchers with genomic and molecular pathways to target through nanotechnological intervention.

Current Therapies

The relatively low incidences of UM and the considerable lack of extensive genomic and proteomic studies of UM tissues contribute to low number of nanomedicines to target UM. However, the versatile nature of nanomedicines suggests its possible successes in treating UM.

Wang et al. investigated the role of dendrimer nanoparticle transfection therapy in the human choroidal melanoma cell line (OCM-1) as a potential treatment (Yingchih Wang, Mo, Wei, & Shi, 2013). The dendrimer nanoparticles were complexed with recombinant DNA plasmids of tumor necrosis factor-a (TNFa) and herpes simplex virus thymidine kinase (HSV-TK), both constructed with the promoter sequence of early growth response-1 (Egr1). Dendrimer nanoparticles are biocompatible and are formed from nanoparticle polymers with sizes of less than 100 nm, physiologically relevant when treating ocular diseases. The polymer is composed of amino groups which are protonated at physiological pH levels, allowing electrostatic interactions with oligonucleotides and their compaction and protection during transfection. The transfection complex is able to be endocytosed by the cell, allowing the oligonucleotide complex to be released into the cell and enter the nucleus for transcriptional regulation. Dendrimer-based nanoparticle transfection complexes offer biocompatibility and the protection of associated nucleic acids (Yingchih Wang et al., 2013).

Molecular studies in various tumors have identified the HSV-TK suicide gene as a promising and successful cancer treatment by killing infected cells and surrounding uninfected cells when administered in conjunction with other drugs or compounds (Rainov, 2000). Wang et al. administered HSV-TK with TNF-a due to the natural anti-tumor properties exhibited by TNF-a, including the ability to activate immune responses and cause hemorrhaging and necrosis in tumoric tissues (Ha Thi et al., 2013; Kearney et al., 2013). HSV-TK and TNF-a were complexed with Egr1 promoter due to Egr1's revealed anti-tumor properties when associated with target genes (Y. Zhou et al., 2010) and the hypothesis that irradiation with ^{125}I may activate the Egr1 promoter complex and induce the transcription and translation of TNF-a and HSV-TK (Yingchih Wang et al., 2013).

Wang et al. radiated the successfully-transfected tissues with ^{125}I and found a significantly elevated expression of TNF-a and HSV-TK through western blot and ELISA analysis after radiation as compared to before, thought to be due to the activation of Egr1 transcription. Transmission electron microscopy (TEM) analysis of irradiated TNF-TK transfected tissues revealed an increased number of cells in the

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

necrotic state along with the inhibition of cell growth and proliferation. The successful coupling of radiation with gene therapy revealed that targeted gene therapy through the Egr1 promoter and radiation could provide for a novel therapeutic route against UM (Yingchih Wang et al., 2013).

Future proteomic studies that reveal a number of other proteins involved in UM pathological pathways may also become similar targets for *in vitro* and *in vivo* experiments such as the one conducted by Wang et al. Coupling nanomedicine gene therapy with standard radiation therapy reveals a synergistic therapeutic effect in treated tissues, and may lead to clinical breakthroughs and better therapies.

Age-Related Macular Degeneration

Pathophysiology

Age-related macular degeneration (AMD) is a disease that destroys sharp, central vision. It affects the central region of the retina known as the macula, which is responsible for fine vision. It is characterized by the presence of drusen and area of hyperpigmentation on the retinal pigment epithelium (RPE). Advanced AMD is characterized by choroidal neovascularization (CNV), the growth of abnormal blood vessels beneath the RPE or between the RPE and the retina(Bylsma & Guymer, 2005). It is accompanied by fluid and blood rupturing Bruch's membrane into the subretinal space, leading to irregularities of the retina.

Current Genomic and Proteomic Information

Case studies have provided evidence for a variety of genetic markers involved in the occurrence of AMD, namely in the formation of drusen and degeneration of macula. Docosahexaenoic acid (DHA) is present in photoreceptor cell membranes in the retina. It is extremely sensitive to oxidative damage and cleavage results in the production of carboxyethylpyrrole (CEP), an oxidative protein (Gu et al., 2010). CEP has been found to be abundant in ocular tissue of patients presenting with AMD. While a genetic marker of oxidative damage, CEPs promote the growth of capillaries and can contribute to neovascularization (Lu et al., 2009).

Due to the neovascularization present in AMD, anti-angiogenic therapy is useful to slow the progression of the disease. Vascular endothelial growth factor A (VEGF-A) is the most potent promoter of angiogenesis and vascular permeability and its role in the pathogenesis of neovascular AMD is well recognized (Ferrara, Gerber, & LeCouter, 2003; Park, Rhu, Kang, & Roh, 2012; Verteporfin In Photodynamic Therapy Study, 2001). VEGF levels and vitreous levels are elevated in human CNV in compared to healthy controls (Sendrowski, Jaanus, Semes, & Stern, 2008). Due to the implication of VEGF in the progression of AMD, anti-angiogenic drugs have been recently pursued to block the development and leakage of new, abnormal blood vessels.

Current Therapies

Three anti-angiogenic drugs are currently being used in the treatment of AMD: ranibizumab (Lucentis), bevacizumab (Avastin), and pegaptanib (Macugen)(Bylsma & Guymer, 2005). Ranibizumab is a human recombinant antibody fragment that displays high binding affinity towards all VEGF isoforms. Clinical trials have shown that ranibizumab helps maintain stable vision without further progression of the disease; however, because of the high cost of the drug, the use of the drug worldwide is limited

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

(Rosenfeld et al., 2006). Bevacizumab has been used more recently as an 'off label' therapy. It is a full-length human recombinant monoclonal antibody, which binds all VEGF isoforms. It is FDA approved for colorectal, lung, and breast cancer, but is used in clinical trials for AMD due to lower patient cost. Pegaptanib is a pegylated aptamer that acts as an anti-VEGF agent. It binds the VEGF165 isoform and inhibits angiogenesis (Vadlapudi, Ashaben, Kishore, & Ashim, 2012). For these anti-angiogenic drugs, the biggest challenge is route of administration. Intravitreal injections allow for the most direct approach; however, the chronic nature of the disease requires consistent injections resulting in side-effects such as retinal detachment and cataract formation (Mudunuri, 2008). Recent studies using nanoemulsions and polymeric micelles containing anti-VEGF bevacizumab result in sustained delivery. Additional research has shown that pDNA encapsulated by micelles can ameliorate choroidal neovascularization in AMD (F. S. Ong et al., 2013). These nanoemulsions and polymeric micelles used for delivery offer more effective drug delivery by their ability to maintain therapeutic concentrations of the active drug molecule over a longer duration than previous methods.

Diabetic Retinopathy

Pathophysiology

Diabetic retinopathy (DR) is a consequence of diabetes, usually developing within 5-10 years. The disease is associated with progressive retinal ischemia (A. F. Clark & T. Yorio, 2003). It is characterized by loss of pericytes in retinal capillaries and subsequent thickening of capillary basement membranes and local ischemia. This induces the expression of VEGF, which can cause vessel leakage and formation of abnormal blood vessels, similar to AMD. The majority of vision loss from DR is due to macular edema. Other factors such as proliferation of fibrovascular membranes can lead to retinal detachment (A. F. Clark & T. Yorio, 2003).

Current Genomic and Proteomic Information

DR has been studied by proteomic analysis of the vitreous humor. Mass spectroscopy based studies have shown that elevated levels of extracellular carbonic anhydrase-I and kallikrein are present in patients with DR. This suggests that retinal hemorrhage contribute to the DR proteomic information. Studies displaying intravitreal injection of carbonic anhydrase-I in rats lead to increased retinal vessel leakage and edema (Gao et al., 2007).

Current Therapies

Current therapies include photocoagulation therapy and pharmacological agents that block VEGF signaling, such as ruboxistaurin mesylate which is a protein kinase C-β inhibitor. Additionally, corticosteroids are also being evaluated for their angiostatic, antipermeable and antifibrotic properties (Samudre, Lattanzio, Williams, & Sheppard, 2004; Y. Wang, Wang, & Chan, 2011). Triamcinolone acetonide and dexamethasone are used in combination with other treatments for posterior ocular disorders. They act by binding steroid receptors in cells to induce or repress targeted genes, thereby inhibiting inflamma-

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

tory symptoms like edema and vascular permeability (Sherif & Pleyer, 2002). Corticosteroids act on VEGF by inhibiting VEGF secretion and inhibiting cytokine production (Wu, Wang, Yang, Huang, & Kuo, 2006). Corticosteroids can also inhibit basic fibroblast growth factor and ICAM-1 expression and decrease VEGF levels that are evident in neovascularization (Penfold et al., 2000; Y. S. Wang, Friedrichs, Eichler, Hoffmann, & Wiedemann, 2002). As with AMD, nanoparticles have gained attention as a drug delivery system to overcome the blood-retinal barriers. Recent studies have encapsulated current therapies in nanoemulsions to sustain drug delivery. Alternatively AuNPs alone have been shown to inhibit retinal neovascularization by suppressing VEGF receptor activation. It is believed that AuNP bind to the heparin-binding proteins of certain VEGF isoforms (Jo, Lee, & Kim, 2011).

CONCLUSION AND FUTURE PERSPECTIVES

Genomic and proteomic analyses consist of a number of different techniques to help identify the molecular basis behind diseases. Because the eyes act as our window to the world around us, measures taken to protect the eyes from damage would be of great benefit to all people.

We have described a number of different ocular diseases, ranging in severity from slight discomfort caused by IOP in glaucomatic patients, to partial blindness in AMD patients, leading to death in patients of UM. We have described a number of molecular markers, genomic and proteomic, that act as indicators for their respective diseases. These genes and proteins are involved in a multitude of pathways in ocular tissues and may play different roles in different diseases, but elucidating their role in ocular diseases helps in developing therapies against those diseases. However, the data obtained from genomic and proteomic analyses in ocular tissues available in current literature is not exhaustive; further research needs to be conducted to identify many more omics-related markers in all kinds of ocular tissues and diseases.

What has been reported so far in regards to genomic and proteomic biomarkers of ocular diseases have been useful in the development of nanomedicine for those diseases. Nanomedicine offers a number of advantages when compared to conventional drug therapy, including small sizes, tissue-targeting, sustained release of the drug, low toxicity, biocompatibility, and biodegradability. Nanomedicine is between 1-999 nm in size, but most are around 100-200 nm in size, sizes that are tolerated in ocular tissues. The sizes and targeting of tissues through surface conjugations of the nanomedicine vectors enable their endocytosis into the cell. Once in the cell, nanomedicine exhibits a steady and sustained release of the drug compound over days, weeks, and even months. Most nanomedicine vectors are made from organic compounds that are well tolerated in ocular tissues, such as PLGA, and can be broken down into harmless compounds by catabolismand eventually excreted. The ability to deliver novel drug, nucleic acid, protein, and herbal compounds through complexing or encapsulating them within nanomedicine allows for the direct targeting of the molecular pathways indicated by omics analysis specific for that disease. Through directly targeting specific proteins, the disease can be prevented, arrested, or even cured at the transcriptional or translational level. Further and continued omics-based research will have a direct impact on the well-being and prognosis of degenerative ocular diseases through the development of novel nanomedicine formulations.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

ACKNOWLEDGMENT

Anjali Hirani and Aditya Grover are equal authors to the manuscript.

REFERENCES

Albery, W.J., & Knowles, J. R. (1976). Free-energy profile for the reaction catalyzed by triosephosphate isomerase. *Biochemistry*, *15*(25), 5627–5631. doi:10.1021/bi00670a031 PMID:999838

Aldrian, S., Trautinger, F., Fröhlich, I., Berger, W., Micksche, M., & Kindas-Mügge, I. (2002). Overexpression of Hsp27 affects the metastatic phenotype of human melanoma cells in vitro. *Cell Stress & Chaperones*, *7*(2), 177. doi:10.1379/1466-1268(2002)007<0177:OOHATM>2.0.CO;2 PMID:12380685

Amendt, B. A., Semina, E. V., & Alward, W. L. M. (2000). Rieger syndrome: A clinical, molecular, and biochemical analysis. *Cellular and Molecular Life Sciences CMLS*, *57*(11), 1652–1666. doi:10.1007/PL00000647 PMID:11092457

Bedikian, A. Y. (2006). Metastatic uveal melanoma therapy: Current options. *International Ophthalmology Clinics*, *46*(1), 151–166. doi:10.1097/01.iio.0000195852.08453.de PMID:16365561

Bhattacharya, S. K., Rockwood, E. J., Smith, S. D., Bonilha, V. L., Crabb, J. S., & Kuchtey, R. W. et al. (2005). Proteomics reveal Cochlin deposits associated with glaucomatous trabecular meshwork. *The Journal of Biological Chemistry*, *280*(7), 6080–6084. doi:10.1074/jbc.M411233200 PMID:15579465

Bui, T., & Thompson, C. B. (2006). Cancer's sweet tooth. *Cancer Cell*, *9*(6), 419–420. doi:10.1016/j.ccr.2006.05.012 PMID:16766260

Bylsma, G. W., & Guymer, R. H. (2005). Treatment of age-related macular degeneration. *Clinical & Experimental Optometry*, *88*(5), 322–334. doi:10.1111/j.1444-0938.2005.tb06716.x PMID:16255691

Chen, G., Gharib, T. G., Huang, C.-C., Thomas, D. G., & Shedden, K. A, Taylor, J. M.G.,... Iannettoni, M. D. (2002). Proteomic analysis of lung adenocarcinoma identification of a highly expressed set of proteins in tumors. *Clinical Cancer Research*, *8*(7), 2298–2305. PMID:12114434

Chu, T.-C., He, Q., & Potter, D. E. (2002). Biodegradable calcium phosphate nanoparticles as a new vehicle for delivery of a potential ocular hypotensive agent. *Journal of Ocular Pharmacology and Therapeutics*, *18*(6), 507–514. doi:10.1089/108076802321021054 PMID:12537677

Clark, A. F., & Yorio, T. (2003). Ophthalmic drug discovery. *Nature Reviews. Drug Discovery*, *2*(6), 448–459. doi:10.1038/nrd1106 PMID:12776220

Clark, A. F., & Yorio, T. (2003). Ophthalmic drug discovery. *Nature Reviews. Drug Discovery*, *2*(6), 448–459. doi:10.1038/nrd1106 PMID:12776220

Coupland, S. E., Vorum, H., Mandal, N., Kalirai, H., Honoré, B., & Urbak, S. F. et al. (2010). Proteomics of uveal melanomas suggests HSP-27 as a possible surrogate marker of chromosome 3 loss. *Investigative Ophthalmology & Visual Science*, *51*(1), 12–20. doi:10.1167/iovs.09-3913 PMID:19643972

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Dutt Vadlapudi, A., Patel, A., Cholkar, K., & K. Mitra, A. (2012, March01). Recent Patents on Emerging Therapeutics for the Treatment of Glaucoma, Age Related Macular Degeneration and Uveitis. *Recent Patents on Biomedical Engineeringe*, *5*(1), 83–101. doi:10.2174/1874764711205010083

Fautsch, M. P., Bahler, C. K., Jewison, D. J., & Johnson, D. H. (2000). Recombinant TIGR/MYOC increases outflow resistance in the human anterior segment. *Investigative Ophthalmology & Visual Science*, *41*(13), 4163–4168. PMID:11095610

Ferrara, N., Gerber, H. P., & LeCouter, J. (2003). The biology of VEGF and its receptors. *Nature Medicine*, *9*(6), 669–676. doi:10.1038/nm0603-669 PMID:12778165

Gambhire, S., Bhalerao, K., & Singh, S. (2013). In situ hydrogel: Different approaches to ocular drug delivery. *Int. J. Pharm. Pharm Sci.*, *5*(2), 27–36.

Gao, B. B., Clermont, A., Rook, S., Fonda, S. J., Srinivasan, V. J., & Wojtkowski, M. et al. (2007). Extracellular carbonic anhydrase mediates hemorrhagic retinal and cerebral vascular permeability through prekallikrein activation. *Nature Medicine*, *13*(2), 181–188. doi:10.1038/nm1534 PMID:17259996

Geroski, D. H., & Edelhauser, H. F. (2000). Drug delivery for posterior segment eye disease. *Investigative Ophthalmology & Visual Science*, *41*(5), 961–964. PMID:10752928

Gomes dos Santos, A. L., Bochot, A., Doyle, A., Tsapis, N., Siepmann, J., & Siepmann, F. et al. (2006). Sustained release of nanosized complexes of polyethylenimine and anti-TGF-β2 oligonucleotide improves the outcome of glaucoma surgery. *Journal of Controlled Release*, *112*(3), 369–381. doi:10.1016/j.jconrel.2006.02.010 PMID:16644054

Gu, J., Pauer, G. J., Yue, X., Narendra, U., Sturgill, G. M., & Bena, J. et al. (2010). Proteomic and genomic biomarkers for age-related macular degeneration. *Advances in Experimental Medicine and Biology*, *664*, 411–417. doi:10.1007/978-1-4419-1399-9_47 PMID:20238042

Ha Thi, H. T., Lim, H.-S., Kim, J., Kim, Y.-M., Kim, H.-Y., & Hong, S.(2013). Transcriptional and post-translational regulation of Bim is essential for TGF-β and TNF-α-induced apoptosis of gastric cancer cell. *General Subjects*, *1830*(6), 3584–3592. doi:10.1016/j.bbagen.2013.03.006

Hashimoto, T., Kusakabe, T., Sugino, T., Fukuda, T., Watanabe, K., & Sato, Y. et al. (2004). Expression of heart-type fatty acid-binding protein in human gastric carcinoma and its association with tumor aggressiveness, metastasis and poor prognosis. *Pathobiology*, *71*(5), 267–273. doi:10.1159/000080061 PMID:15459486

Jiang, C., Moore, M. J., Zhang, X., Klassen, H., Langer, R., & Young, M. (2007). Intravitreal injections of GDNF-loaded biodegradable microspheres are neuroprotective in a rat model of glaucoma. *Molecular Vision*, *13*, 1783–1792. PMID:17960131

Jmor, F., Kalirai, H., Taktak, A., Damato, B., & Coupland, S. E. (2012). HSP-27 protein expression in uveal melanoma: Correlation with predicted survival. *Acta Ophthalmologica*, *90*(6), 534–539. doi:10.1111/j.1755-3768.2010.02038.x PMID:21114636

Jo, D. H., Lee, T. G., & Kim, J. H. (2011). Nanotechnology and nanotoxicology in retinopathy. *International Journal of Molecular Sciences*, *12*(12), 8288–8301. doi:10.3390/ijms12118288 PMID:22174664

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Juronen, E., Tasa, G., Veromann, S., Parts, L., Tiidla, A., & Pulges, R. et al. (2000). Polymorphic glutathione S-transferase M1 is a risk factor of primary open-angle glaucoma among Estonians. *Experimental Eye Research*, *71*(5), 447–452. doi:10.1006/exer.2000.0899 PMID:11040079

Kearney, C. J., Sheridan, C., Cullen, S. P., Tynan, G. A., Logue, S. E., & Afonina, I. S. et al. (2013). Inhibitor of apoptosis proteins (IAPs) and their antagonists regulate spontaneous and tumor necrosis factor (TNF)-induced proinflammatory cytokine and chemokine production. *The Journal of Biological Chemistry*, *288*(7), 4878–4890. doi:10.1074/jbc.M112.422410 PMID:23275336

Kim, J. E., Koo, K. H., Kim, Y. H., Sohn, J., & Park, Y. G. (2008). Identification of potential lung cancer biomarkers using an in vitro carcinogenesis model. *Experimental & Molecular Medicine*, *40*(6), 709–720. doi:10.3858/emm.2008.40.6.709 PMID:19116456

Kostenko, S., & Moens, U. (2009). Heat shock protein 27 phosphorylation: kinases, phosphatases, functions and pathology. *Cellular and Molecular Life Sciences*, *66*(20), 3289–3307. doi:10.1007/s00018-009-0086-3 PMID:19593530

Lee, W.-H., Choi, J.-S., Byun, M.-R., Koo, K.-, Shin, S., Lee, S.-K., & Surh, Y.-J. (2010). Functional inactivation of triosephosphate isomerase through phosphorylation during etoposide-induced apoptosis in HeLa cells: Potential role of Cdk2. *Toxicology*, *278*(2), 224–228. doi:10.1016/j.tox.2010.02.005 PMID:20149834

Lichtenfels, R., Dressler, S. P., Zobawa, M., Recktenwald, C. V., Ackermann, A., & Atkins, D. et al. (2009). Systematic comparative protein expression profiling of clear cell renal cell carcinoma a pilot study based on the separation of tissue specimens by two-dimensional gel electrophoresis. *Molecular & Cellular Proteomics*, *8*(12), 2827–2842. doi:10.1074/mcp.M900168-MCP200 PMID:19752005

Linge, A., Kennedy, S., O'Flynn, D., Beatty, S., Moriarty, P., & Henry, M. et al. (2012). Differential expression of fourteen proteins between uveal melanoma from patients who subsequently developed distant metastases versus those who did not. *Investigative Ophthalmology & Visual Science*, *53*(8), 4634–4643. doi:10.1167/iovs.11-9019 PMID:22570344

Lorigan, J. G., Wallace, S., & Mavligit, G. M. (1991). The prevalence and location of metastases from ocular melanoma: Imaging study in 110 patients. *AJR. American Journal of Roentgenology*, *157*(6), 1279–1281. doi:10.2214/ajr.157.6.1950883 PMID:1950883

Lu, L., Gu, X., Hong, L., Laird, J., Jaffe, K., & Choi, J. et al. (2009). Synthesis and structural characterization of carboxyethylpyrrole-modified proteins: Mediators of age-related macular degeneration. *Bioorganic & Medicinal Chemistry*, *17*(21), 7548–7561. doi:10.1016/j.bmc.2009.09.009 PMID:19786352

Michels-Rautenstrauss, K. G., Mardin, C. Y., Zenker, M., Jordan, N., Gusek-Schneider, G.-C., & Rautenstrauss, B. W. (2001). Primary congenital glaucoma: Three case reports on novel mutations and combinations of mutations in the GLC3A (CYP1B1) gene. *Journal of Glaucoma*, *10*(4), 354–357. doi:10.1097/00061198-200108000-00017 PMID:11558822

Mudunuri, K. (2008). *Intravitreal Delivery of Corticosteroid Nanoparticles.* (Ph.D. Dissertation). University of Florida, Gainesville, FL. Retrieved from http://ufdc.ufl.edu/UFE0022065/00001

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

NEI. (2012). Prevalence of Adult Vision Impairment and Age-Related Eye Diseases in America. *Retrieved*, *11/20*(13), 2013.

Nicolini, C. (2006). Nanogenomics for medicine. *Nanomedicine (Lond)*, *1*(2), 147–152. doi:10.2217/17435889.1.2.147 PMID:17716103

Ong, F. S., Kuo, J. Z., Wu, W. C., Cheng, C. Y., Blackwell, W. L., & Taylor, B. L. et al. (2013). Personalized Medicine in Ophthalmology: From Pharmacogenetic Biomarkers to Therapeutic and Dosage Optimization. *J Pers Med*, *3*(1), 40–69. doi:10.3390/jpm3010040 PMID:24624293

Ong, F. S., Kuo, J. Z., Wu, W.-C., Cheng, C.-Y., Blackwell, W.-L. B., & Taylor, B. L. et al. (2013). Personalized Medicine in Ophthalmology: From Pharmacogenetic Biomarkers to Therapeutic and Dosage Optimization. *Journal of Personalized Medicine*, *3*(1), 40–69. doi:10.3390/jpm3010040 PMID:24624293

Park, Y. G., Rhu, H. W., Kang, S., & Roh, Y. J. (2012). New Approach of Anti-VEGF Agents for Age-Related Macular Degeneration. *Journal of Ophthalmology*, *2012*. doi:10.1155/2012/637316 PMID:22496964

Penfold, P. L., Wen, L., Madigan, M. C., Gillies, M. C., King, N. J., & Provis, J. M. (2000). Triamcinolone acetonide modulates permeability and intercellular adhesion molecule-1 (ICAM-1) expression of the ECV304 cell line: Implications for macular degeneration. *Clinical and Experimental Immunology*, *121*(3), 458–465. doi:10.1046/j.1365-2249.2000.01316.x PMID:10971511

Peyman, G. A., Lad, E. M., & Moshfeghi, D. M. (2009). Intravitreal injection of therapeutic agents. *Retina (Philadelphia, Pa.)*, *29*(7), 875–912. doi:10.1097/IAE.0b013e3181a94f01 PMID:19584648

Qian, Xiao-Long, Shi, Qing-Guo, Pang, Bo, Wu, Rui-Qin, Yu, Lan, Li, Shan-Hu,... Zhou, Jian-Guang. (2010). Identification and expression of two new secretory proteins associated with prostate cancer. *Yi chuan= Hereditas/Zhongguo yi chuan xue hui bian ji, 32*(3), 235-241.

Quigley, H. A. (1996). Number of people with glaucoma worldwide. *The British Journal of Ophthalmology*, *80*(5), 389–393. doi:10.1136/bjo.80.5.389 PMID:8695555

Rainov, N. G. (2000). A phase III clinical evaluation of herpes simplex virus type 1 thymidine kinase and ganciclovir gene therapy as an adjuvant to surgical resection and radiation in adults with previously untreated glioblastoma multiforme. *Human Gene Therapy*, *11*(17), 2389–2401. doi:10.1089/104303400750038499 PMID:11096443

Ramasamy, P., Murphy, C. C., Clynes, M., Horgan, N., Moriarty, P., & Tiernan, D. et al. (2014). Proteomics in uveal melanoma. *Experimental Eye Research*, *118*, 1–12. doi:10.1016/j.exer.2013.09.005 PMID:24056206

Rosenfeld, P. J., Brown, D. M., Heier, J. S., Boyer, D. S., Kaiser, P. K., Chung, C. Y., & Kim, R. Y. (2006). Ranibizumab for neovascular age-related macular degeneration. *The New England Journal of Medicine*, *355*(14), 1419–1431. doi:10.1056/NEJMoa054481 PMID:17021318

Samudre, S. S., Lattanzio, F. A. Jr, Williams, P. B., & Sheppard, J. D. Jr. (2004). Comparison of topical steroids for acute anterior uveitis. *Journal of Ocular Pharmacology and Therapeutics*, *20*(6), 533–547. doi:10.1089/jop.2004.20.533 PMID:15684812

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Selicharova, I., Sanda, M., Miadkova, J., Ohri, S. S., Vashishta, A., Fusek, M.,... Vetvicka, V. (2008). 2-DE analysis of breast cancer cell lines 1833 and 4175 with distinct metastatic organ-specific potentials: comparison with parental cell line MDA-MB-231. *Oncology Reports, 19*(5), 1237-1244.

Sendrowski, D. P., Jaanus, S. D., Semes, L. P., & Stern, M. E. (2008). *Anti-Inflammatory Drugs* (5th ed.). New York, NY: Elsevier Health Sciences.

Sherif, Z., & Pleyer, U. (2002). Corticosteroids in ophthalmology: Past-present-future. *Ophthalmologica, 216*(5), 305-315. doi: 66189

Sigurdsson, H. H., Konráðsdóttir, F., Loftsson, T., & Stefánsson, E. (2007). Topical and systemic absorption in delivery of dexamethasone to the anterior and posterior segments of the eye. *Acta Ophthalmologica Scandinavica*, *85*(6), 598–602. doi:10.1111/j.1600-0420.2007.00885.x PMID:17645424

Singh, A. D., & Borden, E. C. (2005). Metastatic uveal melanoma.[ix.]. *Ophthalmology Clinics of North America*, *18*(1), 143–150. doi:10.1016/j.ohc.2004.07.003 PMID:15763199

Singh, A. D., & Topham, A. (2003). Survival rates with uveal melanoma in the United States: 1973–1997. *Ophthalmology*, *110*(5), 962–965. doi:10.1016/S0161-6420(03)00077-0 PMID:12750098

Steely, H. T. Jr, & Clark, A. F. (2000). The use of proteomics in ophthalmic research. *Pharmacogenomics*, *1*(3), 267–280. doi:10.1517/14622416.1.3.267 PMID:11256578

Stoilov, I., Jansson, I., Sarfarazi, M., & Schenkman, J. B. (2001). Roles of cytochrome p450 in development. *Drug Metabolism and Drug Interactions*, *18*(1), 33–55. doi:10.1515/DMDI.2001.18.1.33 PMID:11522124

Stone, E. M., & Fingert, J. H. (1997). Identification of a gene that causes primary open angle glaucoma. *Science*, *275*(5300), 668–670. doi:10.1126/science.275.5300.668 PMID:9005853

Urtti, A. (2006). Challenges and obstacles of ocular pharmacokinetics and drug delivery. *Advanced Drug Delivery Reviews*, *58*(11), 1131–1135. doi:10.1016/j.addr.2006.07.027 PMID:17097758

Verteporfin In Photodynamic Therapy Study, Group. (2001). Verteporfin therapy of subfoveal choroidal neovascularization in age-related macular degeneration: two-year results of a randomized clinical trial including lesions with occult with no classic choroidal neovascularization--verteporfin in photodynamic therapy report 2. *Am J Ophthalmol, 131*(5), 541-560.

Wadhwa, S., Paliwal, R., Paliwal, S. R., & Vyas, S. P. (2010). Hyaluronic acid modified chitosan nanoparticles for effective management of glaucoma: development, characterization, and evaluation. *Journal of Drug Targeting*, *18*(4), 292–302. doi:10.3109/10611860903450023 PMID:19943753

Wang, , Wei, W. B., Mo, L., & Shi, X.-H. (2013). Efficacy and safety of dendrimer nanoparticles with coexpression of tumor necrosis factor-α and herpes simplex virus thymidine kinase in gene radiotherapy of the human uveal melanoma OCM-1 cell line. *International Journal of Nanomedicine*, *8*, 3805. doi:10.2147/IJN.S48950 PMID:24124368

Wang, Y., Wang, V. M., & Chan, C. C. (2011). The role of anti-inflammatory agents in age-related macular degeneration (AMD) treatment. *Eye (London, England)*, *25*(2), 127–139. doi:10.1038/eye.2010.196 PMID:21183941

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Wang, Y. S., Friedrichs, U., Eichler, W., Hoffmann, S., & Wiedemann, P. (2002). Inhibitory effects of triamcinolone acetonide on bFGF-induced migration and tube formation in choroidal microvascular endothelial cells. *Graefe's archive for clinical and experimental ophthalmology = Albrecht von Graefes Archiv für klinische und experimentelle Ophthalmologie*, *240*(1), 42–48. doi:10.1007/s00417-001-0398-y PMID:11954780

Weinreb, R. N., Kashiwagi, K., Kashiwagi, F., Tsukahara, S., & Lindsey, J. D. (1997). Prostaglandins increase matrix metalloproteinase release from human ciliary smooth muscle cells. *Investigative Ophthalmology & Visual Science*, *38*(13), 2772–2780. PMID:9418730

Wong, T. T.L., Sethi, C., Daniels, J. T., Limb, G. A., Murphy, G., & Khaw, P. T. (2002). Matrix metalloproteinases in disease and repair processes in the anterior segment. *Survey of Ophthalmology, 47*(3), 239-256.

Wu, W. S., Wang, F. S., Yang, K. D., Huang, C. C., & Kuo, Y. R. (2006). Dexamethasone induction of keloid regression through effective suppression of VEGF expression and keloid fibroblast proliferation. *The Journal of Investigative Dermatology*, *126*(6), 1264–1271. doi:10.1038/sj.jid.5700274 PMID:16575391

WuDunn, D.WuDunn. (2002). Genetic basis of glaucoma. *Current Opinion in Ophthalmology*, *13*(2), 55–60. doi:10.1097/00055735-200204000-00001 PMID:11880716

Zarbin, M. A., Montemagno, C., Leary, J. F., & Ritch, R. (2013). Nanomedicine for the treatment of retinal and optic nerve diseases. *Current Opinion in Pharmacology*, *13*(1), 134–148. doi:10.1016/j.coph.2012.10.003 PMID:23142104

Zhang, L., Cilley, R. E., & Chinoy, M. R. (2000). Suppression subtractive hybridization to identify gene expressions in variant and classic small cell lung cancer cell lines. *The Journal of Surgical Research*, *93*(1), 108–119. doi:10.1006/jsre.2000.5957 PMID:10945951

Zhang, W., Prausnitz, M. R., & Edwards, A. (2004). Model of transient drug diffusion across cornea. *Journal of Controlled Release*, *99*(2), 241–258. doi:10.1016/j.jconrel.2004.07.001 PMID:15380634

Zhao, X., Ramsey, K. E., Stephan, D. A., & Russell, P. (2004). Gene and protein expression changes in human trabecular meshwork cells treated with transforming growth factor-β. *Investigative Ophthalmology & Visual Science*, *45*(11), 4023–4034. doi:10.1167/iovs.04-0535 PMID:15505052

Zhou, Y., Song, X., Jia, R., Wang, H., Dai, L., Xu, X.,... Fan, X. (2010). Radiation-inducible human tumor necrosis factor-related apoptosis-inducing ligand (TRAIL) gene therapy: a novel treatment for radioresistant uveal melanoma. *Pigment Cell & Melanoma Research, 23*(5), 661-674.

Zhou, Z., & Vollrath, D. (1999). A cellular assay distinguishes normal and mutant TIGR/myocilin protein. *Human Molecular Genetics*, *8*(12), 2221–2228. doi:10.1093/hmg/8.12.2221 PMID:10545602

KEY TERMS AND DEFINITIONS

Age-Related Macular Degeneration: Ocular disease that usually affects older adults and results in loss of central, focused vision.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Diabetic Retinopathy: Damage to retina caused by diabetes and can lead to blindness.
Glaucoma: Eye condition that leads to damage of optic nerve due to intraocular pressure.
Nanogenomics: Nanotechnology utilized in the study of genomics.
Nanoproteomics: Nanotechnology utilized in the study of proteomics.
Nanotechnology: Engineering of functional systems at the molecular scale.
Omics: Refers to field of study such as proteomics and genomics.
Uveal Melanoma: Cancer of the eye involving the uvea.

This work was previously published in the Handbook of Research on Diverse Applications of Nanotechnology in Biomedicine, Chemistry, and Engineering edited by Shivani Soni, Amandeep Salhotra, and Mrutyunjay Suar, pages 152-166, copyright year 2015 by Engineering Science Reference (an imprint of IGI Global).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 14
Nanomedicine Magic Bullet for Human Cancer

Srinivas Patnaik
University of Houston, USA

ABSTRACT

Nanotechnology is the new tool that has changed healthcare, engineering, and space science. The technology involves nanoparticles that are effectively a bridge between bulk materials and atomic or molecular structures. The properties of materials change its surface plasmon resonance in metals, supermagnetism in magnetic materials as their size approaches to nanoscale. Taking in to account of their small sizes (less than 100nm) and their miraculous properties, unlike their precursor bulk material, nanoparticles are exploited to create new diagnostics and therapeutics with respect to several human diseases. Nanomedicine is generating a new generation of innovative revolution in nanoscale drug delivery strategies, site-specific drug delivery, and personalized therapy in cancer by releasing the drug at a specific site. This chapter discusses the evolution of nanomedicine to several advancements in the field of nanoparticle technologies, targeting and controlled release strategies, with the desire of generating robust and efficient nanotherapeutic tools against cancer.

INTRODUCTION

Cancer became the greatest cause of considerable morbidity and mortality worldwide, surpassing heart disease as the leading cause of death in the US population (Jemal, Siegel, Xu, & Ward, 2010). The prevalent disconcerting is the fact that several significant established methodologies towards the treatment of the disease have failed and have strongly impacted patient survival. Different ways of cancer treatment in last several decades witnessed the chemotherapy, adjuvant therapeutic as the major tools to tackle cancer. Cell and molecular biology research of human cancer improved understanding of the disease initiation and progression and development of highly specific agents capable of exerting their effects on individual proteins or pathways either over expressed or aberrant within tumors. The specificity of these agents in killing the cancer cells were meant to circumvent toxicities associated with more traditional chemotherapeutics, while resulting in improved high-end therapeutics. Although these novel

DOI: 10.4018/978-1-5225-0549-5.ch014

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

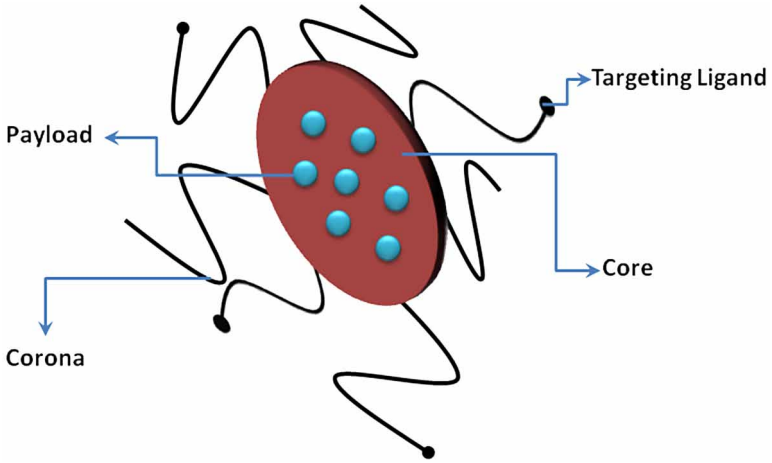
chemotherapeutics have led to improvements in survival, they are still weighed down by a number of biological barriers that deter efficacious drug delivery following systemic administration. Either newly discovered or a traditional chemotherapeutic drug suffers from non-specific relocation, with only a tiny proportion of drugs reaching the tumor. For example intravenously injected chemotherapeutic agents experience sequestration as a result of clearing foreign materials by monocytes and macrophages (Peer et al., 2007). This results in accumulation of drugs in healthy noncancerous tissues, causing severe morbidity. In case of doxorubicin, a drug used in cancer chemotherapy and like all anthracyclines, it works by intercalating DNA, has the most serious adverse effect being life-threatening such as heart damage (Olson & Mushlin, 1990). Even though there are several chemotherapeutics agents effectively kill cancer cell, specificity in targeting tumor disqualify the therapeutic potential of these drugs. Hence a more effective way to deliver them to tumors became the need of the hour.

Nanotechnology is broadly defined as the science and engineering involved in the design, synthesis, characterization, and application of materials and devices with at least one of the dimensions on the nanoscale (typically 1–100 nanometers) (Alexis et al., 2008) It is a multidisciplinary field that uses principles from chemistry, biology, physics, and engineering to design and fabricate nanoscale devices (Farokhzad & Langer, 2006); (Ferrari, 2005); (Fox, 2000); (Jiang, Kim, Rutka, & Chan, 2007); (Peppas, 2004); (Sinha, Kim, Nie, & Shin, 2006); (Uchegbu, 2006). Nanotechnology commonly based on materials up to several hundred nanometers that are developed using top-down or bottom-up engineering. Various engineered nanomaterial reveal distinctive capabilities based on intrinsic properties such as shape, size and unique functional properties conferred through surface modifications. As illustrated in Figure 1, basic components of nanoparticle platforms for drug delivery consists of a core, corona, payload and targeting ligand. Since its origin decades ago, nanotechnology has drawn increasing attention in the field of medical sciences for its requirement in disease diagnostics, prevention, and treatment of human diseases. New discoveries in the field of medicine have encouraged even more for the advances in nanotechnology. Currently there are many cancer therapies in the market or under development based on nanotechnology (Davis, Chen, & Shin, 2008); (Euliss, DuPont, Gratton, & DeSimone, 2006); (Farokhzad, 2008); (Freitas, 2005); (K. K. Jain, 2008); (Kawasaki & Player, 2005); (Lanza et al., 2006); (Levy-Nissenbaum, Radovic-Moreno, Wang, Langer, & Farokhzad, 2008); (Moghimi, Hunter, & Murray, 2005); (Peer, et al., 2007); (Pridgen, Langer, & Farokhzad, 2007); (Riehemann et al., 2009);(Rosen & Abribat, 2005); (Venugopal et al., 2008).

Recently knowing the importance of nanotechnology many developed and developing countries have come forward to make paradigm-changing impacts on the detection, treatment, and prevention of cancer. In fact level of interest in nanotechnology by both academic and industrial investigators has led to increased development of novel nanotechnology platforms for medical applications resulting in sharp increase in government funding, and venture capital investment. Nanotechnology has the potential to make a significant impact in site specific drug delivery (Pridgen, et al., 2007)). Several nanoscale drug delivery systems launched in clinics for diagnostics and therapy, however full blown therapeutic systems are in the advanced stage of development. Nanoscale drug delivery vehicles have shown the ability to encapsulate a variety of therapeutic agents such as small molecules, peptides, protein-based drugs, and nucleic acids. By encapsulating these molecules inside a specialized material known as nanocarrier, the solubility and stability of the drugs can be improved, providing an opportunity to reassess potential drugs previously overlooked because of poor performance in the respective application (Langer, 1998). Encapsulated molecules can be released from nanocarriers in a controlled manner over time to maintain a drug concentration within a therapeutic window or the release can be

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 1. Schematic representation: Basic components of nanoparticle platforms for drug delivery, including core, corona, payload, and targeting ligand



triggered by some stimulus unique to the delivery site (Moghimi, 2006). Nanocarrier can be surface engineered to increase the blood circulation efficiency and optimized distribution, while attachment of targeting ligands to the surface can result in enhanced uptake by target tissues (Moghimi, Hunter, & Murray, 2001). The small size of the nanocarriers escapes from biological barriers and accomplish cellular uptake (Brigger, Dubernet, & Couvreur, 2002). The overall objective of these nanocarriers is to lower the systemic toxicity of the therapeutic agent while increasing the concentration of the agent in the area of interest, resulting in a higher therapeutic index for the therapeutic agent. For detection of tumors imaging agents can also incorporated into nanocarriers to improve tumor detection and imaging (K. Kim et al., 2006); (Montet, Weissleder, & Josephson, 2006). Nanoparticles (NP) can be engineered to be multifunctional with the ability to target diseased tissue, carry imaging agents for detection, and deliver multiple therapeutic agents for combination therapy (Nasongkla et al., 2006). Recently the potential of nanomedicine for addressing cancer treatment were assessed from clinical trials. To name a few, first phase I clinical trial in human patients with solid cancers of a systemically administered targeted nanoparticle containing siRNA was tested in 2008 (Davis, et al., 2008); (Davis et al., 2010). Nanoparticles consisting of a linear, cyclodextrin-based polymer (CDP), transferrin (Tf) protein targeting ligands, and anti-R2 siRNA were shown to accumulate in tumor cells at therapeutically relevant doses. The mRNA cleavage at the site of RNA interference, demonstrates specific gene inhibition by nanotechnology based siRNA system in humans. In this chapter, we will focus on nanoparticle technologies, specifically nanocarrier drug delivery systems like polymeric nanoparticles, dendrimers, nanoshells, liposomes, inorganic/metallic nanoparticles, hybrid nanoparticles, micelles, and magnetic and bacterial nanoparticles for cancer therapy applications. This chapter will include a discussion of how improvements in the understanding of the tumor microenvironment have guided the design of both non-targeted and targeted nanocarriers as therapeutic vehicles for cancer (Bierie & Moses, 2006); (Bissell & Labarge, 2005); (Cairns, Papandreou, & Denko, 2006); (Zetter, 2008). We also aim to highlight similar innovative nanomedicine strategies, focusing on platforms and novel trends that are marshalling the next generation of chemotherapeutic formulations.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

NANOPARTICLE TECHNOLOGIES

Passive Targeting

The observation that certain macromolecules display tumoritropic accumulation, was witnessed more than 30 years ago (Maeda, 2001). Poly(Styrene-co-Maleic Acid)-Neo Carzino Statin (SMANCS), a 16 kDa polymer conjugate that non-covalently binds albumin in the circulation to reach a molecular weight of around 80 kDa reported to be the first tumor accumulation of therapeutic macromolecules (Matsumura 1986). Using labeled albumin and other proteins in addition to the polymer conjugate, they showed that proteins larger than 30 kDa could preferentially distribute to the tumor and remain there for prolonged periods of time (Matsumura 1986). This preferential distribution to the tumors was recognized to the presence of fenestrations in the imperfect tumor blood vessels and to the poor lymphatic drainage in the tissue. The combination of these two phenomena was coined as the enhanced permeation and retention effect. There after it became the favorite tool for scientists for the efficient delivery of anticancer drugs to tumors, whether using polymer conjugates, liposome or NPs. Nonetheless, the EPR effect is much more complex than initially defined and the phenomenon has somehow become a blanket term encompassing dozens of complex biological processes like angiogenesis, vascular permeability, hemodynamic regulation, heterogeneities in tumor genetic profile, heterogeneities in the tumor microenvironment and lymphangiogenesis. These factors vary among patients and their tumor types. Likewise, the distribution and accumulation of NPs in tumors are also multifaceted and are affected by the biological and physicochemical properties of each material. For these reasons, and with the advent of personalized medicine, the designation of EPR as a simple self explaining phenomenon might be becoming outdated.

When a solid tumor reaches a given size, the normal vasculature present in its vicinity is not sufficient to provide all the oxygen supply required for its further proliferation. As cells start to die, they secrete growth factors that trigger the budding of new blood vessels from the surrounding capillaries (Bates, Hillman, Williams, Neal, & Pocock, 2002). This process, known as angiogenesis, promotes the rapid development of new, irregular blood vessels that present a discontinuous epithelium and lack the basal membrane of normal vascular structures (R. K. Jain, 1998); (R. K. Jain & Stylianopoulos, 2010). The resulting fenestrations in the capillaries can reach sizes ranging from 200 to 2000 nm, depending on the tumor type, its environment and its localization (Hobbs et al., 1998). When blood components reach the abnormal, discontinuous vascular bed, the fenestrations offer little resistance to extravasation to the tumor interstitium. This denotes the enhanced permeation portion of the EPR effect. In normal tissues, the extracellular fluid is constantly drained to the lymphatic vessels at a mean flow velocity around 0.1–2 μm/s (Swartz & Fleury, 2007). This allows the continuous draining and renewal of interstitial fluid and the recycling of extravasated solutes and colloids back to the circulation. In tumors, the lymphatic function is defective, resulting in minimal uptake of the interstitial fluid (Padera et al., 2004); the colloids cannot rely on convective forces to return to circulation.

In a solid tumor, the distribution of molecules to the tumor is governed by few distinctive phenomena: the extravasation of colloids from the blood vessels, their further diffusion through the extravascular tissue and their interaction with intracellular and/or extracellular targets within the tumor environment. The first two aspects are the result of diffusive forces and can be influenced concurrently by the tumor biology and the characteristics of the diffusing species. The third parameter is more complex and less understood. It represents the interactions of the colloids with the tumor whether through adsorption phenomena, cellular uptake or degradation and metabolism. These aspects can all affect the equilibrium of

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

accumulation inside the tumor; they depend on the nature of the material, its affinity for all components of the tissue and the tumor composition (e.g., nature of the extracellular matrix, the type of cells present). The tumor environment is a complex situation where normal anatomy and physiology principles are defied. The untamed growth of cancer cells and the secretion of angiogenic factors both contribute to the highly disorganized vasculature and congested extravascular environments. Paradoxically, these structural imperfections are both the cause of the EPR effect and the principal source of uneven tumor accumulation and retention of nanomaterials. The release of angiogenic factors from cancer cells dictates the blood vessel architecture and morphology of the vascular wall (Bates, et al., 2002). The neovasculature sprouting from capillaries is disordered, discontinuous and highly fenestrated (Carmeliet & Jain, 2011). The degree of leakiness of the endothelium and the enhanced vascular permeability to macromolecules and nanomaterials depend on many factors like the cancer type, its stage and in xenograft models, the implantation site of the tumor (R. K. Jain & Stylianopoulos, 2010); (Hobbs, et al., 1998); (Hashizume et al., 2000).

The shape of NPs has been shown to influence blood exposure by modulating interactions with the mononuclear phagocyte system (MPS) (Bertrand & Leroux, 2012); (Geng et al., 2007). Furthermore, single-wall carbon nanotubes with high aspect ratio from 100:1 to 500:1 and were shown to be cleared efficiently by the kidneys, suggesting that elongated shapes could provide benefits to the filtration process through porous structures (Ruggiero et al., 2010). Tumor distribution kinetics of nanorods with a length of 44 nm (aspect ratio: 10) were compared to those of 35-nm nanospheres showing the same hydrodynamic radius (Chauhan et al., 2011). Despite similar blood circulation profiles, the nanorods were shown to extravasate to the interstitium 4 times faster and to diffuse deeper in the tumor, interacting with a 1.7-fold larger volume. Similarly, regardless of their increased splenic clearance and shorter blood circulation times, elongated viral nanofilaments derived from plant viruses showed increased tumor homing and accumulation compared to other spherical viral constructs (Shukla et al., 2013). All together, these findings suggest that elongated architecture might present beneficial EPR properties.

Active Targeting

Ligand-mediated targeting, also known as active targeting involves utilizing affinity ligands on the surface of NPs for specific retention and uptake by the disease cells targeted. To that end, ligands are selected to target surface molecules or receptors overexpressed in diseased organs, tissues, cells or subcellular domains(Peer, et al., 2007); (Kamaly, Xiao, Valencia, Radovic-Moreno, & Farokhzad, 2012); (Koshkaryev, Sawant, Deshpande, & Torchilin, 2013; Shi, Xiao, Kamaly, & Farokhzad, 2011). Actively targeted material needs to be in the proximity of their target to benefit from this increased affinity. Therefore, the approach is aimed toward increasing interactions between NPs and cell and enhancing internalization of drugs without altering the overall bio-distribution(Alexis, Pridgen, Molnar, & Farokhzad, 2008). The design of actively-targeted NP drug carriers is complex because the NP architecture, the ligand conjugation chemistry and the types of ligand available all contribute to the efficacy of the system. Other factors like the administration route or the non-specific binding of proteins during the NP's journey through the bloodstream have been shown to affect the targeting ability of NPs (Gu et al., 2008). Physicochemical properties like the ligand density, the size of the NPs (Jiang, et al., 2007) or the choice of the targeting ligand (Valencia et al., 2011) might also possibly affect the efficacy of the active targeting strategy in vitro and, most importantly in vivo. The following section will highlight the strategies, benefits and drawbacks of combining targeting ligands with NP drug delivery systems in the targeting of solid tumors.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

The main mechanism behind active targeting is the recognition of the ligand by its target substrate. Representative ligands include antibodies, proteins, peptides, nucleic acids, sugars, and small molecules such as vitamins (Saha, Vasanthakumar, Bende, & Snehalatha, 2010). Target molecules can be proteins, sugars or lipids present in diseased organs or on the surface of cells (Yu, Tai, Xue, Lee, & Lee, 2010); (des Rieux, Pourcelle, Cani, Marchand-Brynaert, & Preat, 2013). The interactions of ligand-functionalized NP systems with their target antigen are enhanced by the multivalent nature of the NP architecture and multiple copies of the ligand increase the avidity of the NP for its target (Wang, Tian, Petros, Napier, & Desimone, 2010). The targeting specificity and the delivering capacity are two important aspects to evaluate the efficiency of an active targeting system. The specificity is determined by the bio distribution of the ligand-functionalized NP and by how the conjugated ligand and the NP system interact with off-target molecules and cells; it is defined by the ligand and NP properties. The delivering capacity is directly related to the NP material and structure (Kamaly, et al., 2012). Currently, actively targeted NPs are envisioned as a promising complementary strategy to EPR to further augment the efficiency of cancer nanomedicines. Actively-targeted NPs require being in the vicinity of their target antigen to recognize and interact with it. That intrinsic characteristic is considered a major challenge to the development of actively-targeted NPs(Florence, 2012). Here again, the systemic clearance of the NPs affects the amounts available in the bloodstream supplying the tumor.

Liposome Nanoparticles

Lipids form nanoparticle vesicles through the self-assembly of amphiphilic lipids and excipients. The lipids form a bilayer based on hydrophobic interactions in continuous parallel packing, with the hydrophilic head groups positioned towards the aqueous environment. Hydrophilic molecules can be encapsulated in the inner aqueous phase while hydrophobic molecules can be carried in the hydrophobic domains of the lipid bilayer. Physicochemical properties of liposomes can be precisely changed to control surface charge, functionality, and size by simply mixing commercially available lipid molecules. This offers a significant advantage over other carriers that require much more controlled synthesis steps and additional chemical modifications. Generally, lipids used to prepare vesicular formulations are found in the human body and approved by the FDA, such as DSPE (1, 2-distearoylsn- glycero-3-phosphoethanolamine), HSPC (hydrogenated phosphatidylcholine from soybean lecithin), EggPG (egg yolk phosphatidylglycerol) and DSPC (1, 2-distearoyl-glycero-3-phosphocholine). Each of these lipids can be obtained with or without PEG, which can be used to modify the surface of the resulting liposome. Doxil, a pegylated liposome clinically used to treat multiple types of cancer, is a landmark for liposomal drug delivery systems. Doxil consists of a packed pegylated surface (2 kDa PEG chains) and is loaded with doxorubicin through drug diffusion based on an ammonium salt gradient. This method achieves a stable drug entrapment in a crystal form with reduced leakage over a long period of time. Doxil liposomes have a size of ~100 nm, surface charge of ~10 mV, and a long-term shelf stability of ~2 years at ~4^0C. Recently, Aphios Corp. developed nanosomes (small liposomes, <100 nm) carrying multiple drugs such as docetaxel, camptothecin, bryostatin-1 and vitamin D analogue for treatment of multiple cancer types (Castor, 2005) using a manufacturing technology based on a super-critical fluid process. In addition, Novosom AG uses amphoteric liposomes to deliver nucleic acids. The liposomal formulation is able to change surface charge properties (zeta potential) with changes in solution pH. The charge switch at acidic pH results in fusion with the cell membrane during endocytosis uptake, allowing escape of the nanocarriers into the cytoplasm to deliver the therapeutic load. Liposomal formulations have demonstrated

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

multiple benefits as drug delivery vehicles. However, they must be used to carry very potent drugs due to their low encapsulated load. Lipid-based vesicles pose several other challenges such as instability in the bloodstream, poor solubility of many drugs in the lipid/surfactant solution, and a rapid, burst release of drug. Liposomal formulations are also associated with severe side effects due to their accumulation in skin tissue. While prolonged drug release kinetics are difficult to control using liposomal systems, alternatives such as environmentally triggered release can be easily engineered by inserting destabilizing lipids with amine head groups into the vesicle membrane or including additives such as morpholine in the lipid formulation (Cullis & Chonn, 1998); (Guo, MacKay, & Szoka, 2003); (Kocer, 2007); (Sudimack, Guo, Tjarks, & Lee, 2002); (Vial, Rabhi, & Tribet, 2005). There are currently no liposomal formulations with triggered drug release approved for clinical use or in early phases of clinical trials. However, LiPlasome Pharma developed non-targeted liposomes consisting of lipids designed to be degraded by phospholipase A2 (PLA2), which is up-regulated in the tumor microenvironment (Andresen, Davidsen, Begtrup, Mouritsen, & Jorgensen, 2004); (Jensen et al., 2004); (Jorgensen, Davidsen, & Mouritsen, 2002). The lipid degradation products are converted into anticancer drugs, resulting in local delivery of cytotoxic drugs in the tumor. In-vivo results showed a delay in colon cancer progression using a human tumor xenograft mice model (Tribler et al., 2007). This approach also provides the possibility of multi-drug delivery. Protein stabilization of liposomes is being investigated by Azaya Therapeutics to deliver hydrophobic drugs such as docetaxel for cancer therapy. Docetaxel is encapsulated into the liposome bi-layer and stabilized by albumin to prevent rapid drug leakage (ATI- 1123). The results of ATI-1123 efficacy studies in human xenograft mice models for prostate, pancreatic, and non-small-cell lung cancer (NSCL cancer) showed partial tumor regression in 90% of the PC3 tumor xenograft model and improved efficacy in the pancreas model when compared to groups treated with docetaxel at equal doses (25 mg/ kg). This may be explained by the slower plasma elimination and higher bioavailability of ATI-1123 relative to free docetaxel (Zamboni, 2008).

Polymer–Drug Conjugates Nanoparticles

Polymer–drug conjugates are one of the most investigated types of nanocarriers and is currently in clinical trials as advanced as phase III. Polymer–drug conjugates are formed through side-chain grafting of drugs to polymer chains, allowing them to deliver high doses of chemotherapeutic drugs. Although the physicochemical properties of a number of formulations are not disclosed, the size of polymer–drug conjugates is generally below 20 nm. HPMA-doxorubicin (N-(2-hydroxypropyl) methacrylamide) copolymer (PK1) was the first synthetic polymer–anticancer drug conjugate to enter clinical trials more than a decade ago and the clinical phase II trial for women with advanced breast cancer is still ongoing (Vasey et al., 1999). Similarly, Prolindac (AP5346) is composed of a HPMA backbone copolymer with platinum grafted to the side chains through a pH-sensitive chelator designed for drug release in the tumor environment (Sood et al., 2006). Preclinical data shows superior efficacy of the polymer–drug conjugates using multiple cancer models including a M5076 sarcoma platinum-resistant tumor xenograft mice model, multiple colon xenograft models, L1210 leukemia, and 0157 hybridoma models (Rice, Gerberich, Nowotnik, & Howell, 2006). Oxaliplatin drug loading was ~10% (w/w) using a polymer chain of 25 kDa and the drug release was slow. Formulations were injected once a week for three weeks and the polymer–drug conjugates significantly retarded tumor growth over one month due to higher intracellular concentration of platinum (Pt). In the clinical phase I trial conducted in Europe (Campone et al., 2007), systemic injection of 640 mg Pt /m ^{2} weekly for 3 weeks resulted in a response in platinum-resistant

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

ovarian cancer. Recently, Access Pharmaceuticals Inc. reported the results of the clinical phase II trial showing that 66% of the patients with ovarian cancer experienced meaningful disease stabilization and limited side effects. Polyamino acids grafted with drugs on the side chains are another class of polymer–drug conjugates that have demonstrated high drug loading and efficacy(Matsumura, 2008). In the case of polyglutamate-glycine-campthotecin (CT-2106), degradable linkers have allowed drug loadings ranging from 5% to 50%. Using a glycine linker, drug loadings were increased threefold over polyglutamate- campthotecin alone due to reduced steric hindrance. However, a formulation with a drug load of ~30% was selected for clinical trials due to superior stability and efficacy in human tumor xenograft mice models (Homsi et al., 2007). Meanwhile, Xyotax, a similar polymer–drug conjugate (polyglutamate-paclitaxel), is in 22 clinical trials at the moment for multiple cancer therapies including prostate cancer, metastatic breast cancer, neck cancer, metastatic colorectal cancer, and recurrent NSCL (Phase III). Paclitaxel is grafted to polyglutamic acid (30–40 kDa) to reach a drug load of 20–40% by weight (Singer, 2005; Singer et al., 2003). The clinical data shows an average improvement in median survival in Xyotax patients compared with the control group, although there were no differences in the overall survival. One benefit of the treatment was the reduction of multiple side effects including neurotoxicity (Boddy et al., 2005). Overall, polymer–drug conjugates are considered simple nanocarrier systems, but tuning the optimal formulation might require extensive development. For example, small changes in the polymer–drug conjugation efficiency may significantly modify the pharmacokinetic parameters and tissue biodistribution. The resulting formulation could also be considered a new chemical entity, complicating regulatory approval.

Polymeric Nanoparticles

Polymeric nanoparticles may represent the most effective nanocarriers for prolonged drug delivery. The early in vitro and in vivo development of polymeric nanoparticles loaded with drugs in the 1980s using polyalkylcyanoacrylate-based nanoparticles releasing doxorubicin (Couvreur, Kante, Roland, & Speiser, 1979) led to multiple reports using polymer-based materials for drug delivery. Langer and Folkman (Langer & Folkman, 1976) demonstrated the first controlled release of macromolecules using polymers, which allowed the development of anti-angiogenic drug delivery systems for cancer therapy and opened new areas for the delivery of macromolecules. In 1994, Langer et al. described nanoparticles composed of poly(lactic acid)/poly (lactic-co-glycolic acid) (PLA/PLGA) and PEG block copolymer as “long-circulating nanoparticles” due to their stealth properties (Gref et al., 2000), leading to an increased interest in polymeric nanoparticles and their therapeutic applications. Only a few papers per year were published using polymeric nanoparticles as a drug delivery system in the 1990s in contrast to ~200 papers in 2008. Polymeric nanoparticles provide significant flexibility in design because polymers can be biodegradable or nonbiodegradable, and can be made synthetically or derived from natural sources. Some common polymers used for nanoparticle formation include poly (lactic acid) (PLA), dextran, and chitosan. Biodegradable polymers are typically degraded into individual monomers, which are metabolized and removed from the body via normal metabolic pathways. Degradation and drug release kinetics can be precisely controlled by the physicochemical properties of the polymer, such as molecular weight, dispersity index, hydrophobicity, and crystallinity. In general, drugs can be released in a controlled manner with first-order kinetics due to drug diffusion through the polymeric matrix or triggered in response to the local environment. The nanoparticle surface is usually sterically stabilized by

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

grafting, conjugating, or adsorbing hydrophilic polymers such as PEG to its surface, which can also reduce hepatic uptake and improve circulation half-life (Gref, et al., 2000); (Peracchia et al., 1999). Several polymeric nanoparticles are now in various stages of preclinical and clinical development. For example, Nanolymf Ltd. developed microparticles carrying encapsulated nanocapsules loaded with drugs. Drug-loaded polymethacrylate nanocapsules (~400 nm) are encapsulated in 2–10 mm cellulose-based microspheres and given orally, resulting in uptake by M-cells and a drug blood bioavailability of ~5%. DeSimone et al. (Euliss, et al., 2006; Gratton, Napier, Ropp, Tian, & DeSimone, 2008; Gratton et al., 2007; Kelly & DeSimone, 2008; Rolland et al., 2005) have shown that physicochemical properties of particles such as shape, size and mechanical flexibility contribute to their interactions with cell membranes and control their internalization pathways. This has led to the preclinical development of polymeric nanoparticles using a "PRINT" technology (Particle Replication In Non-wetting Templates) for cancer therapy and other diseases.

Micelle Nanoparticles

Micelles are composed of lipids or other amphiphilic molecules, such as polymers or polyamino acids, and self-assemble into small nanoparticles composed of a hydrophobic core. Micelles have been developed as drug delivery carriers for hydrophobic drugs (Aliabadi, Shahin, Brocks, & Lavasanifar, 2008); (Liggins & Burt, 2002); (Matsumura, 2008). There are multiple examples of micellar formulations under investigation or in clinical trials, such as Genexol-PM (D. W. Kim et al., 2007); (K. S. Lee et al., 2008), NC-6004 (Uchino et al., 2005), NK105 (Hamaguchi et al., 2007), and NK911 (Matsumura et al., 2004); (Tsukioka et al., 2002). Genexol-PM is the first non targeted polymeric micellar formulation approved for cancer therapy. It was approved in Korea in 2006 as a first-line therapy for metastatic breast and non-small cell lung cancer (NSCL) cancer (currently in Phase III). It is currently being evaluated in a clinical phase II trial in the USA for metastatic pancreatic cancer therapy. Genexol-PM is composed of a block copolymer PDLLA (1.75 kDa)–mPEG (2 kDa) forming micelles with a size of ~60 nm and paclitaxel loading of ~15% (w/w). The maximum tolerated dose (MTD) of Genexol-PM is threefold higher than Taxol (60 mg/kg vs. 20 mg/kg, respectively) and the median lethal tolerated dose (LD50) using Sprague–Dawley rats was reported to be ~20 times higher than Taxol. Interestingly, the area under the plasma concentration (AUC) was similar for both formulations. However, paclitaxel had more significant accumulation in tissues such as the liver and tumor with the Genexol-PM formulation, leading to differential tumor cytotoxicity and reduction of tumor volume (S. C. Kim et al., 2001). Results of a clinical phase I trial showed that while the MTD was almost double (390 mg/ m^2) for Genexol-PM compared to Taxol with similar toxicological profiles, the recommended dose was determined to be 300 mg/m^2 (T. Y. Kim et al., 2004). The clinical phase II trial in Korea evaluated Genexol-PM as a co-therapy with cisplatin for advanced NSCL in contrast to a single agent therapy (D. W. Kim, et al., 2007). The clinical phase II results showed ~30% of the patients had stable disease status and 60% of the patients had an increased survival of one year using slightly lower doses of cisplatin than with the combined treatment of Taxol with cisplatin (60 mg/ m^2 versus 75 mg/ m^2, respectively) (D. W. Kim, et al., 2007). Other companies such as Labopharm and Intezym are also developing micelle systems for the delivery of a myriad of anticancer agents using formulations with sizes ranging from 10 to 200 nm using polyamino acids and synthetic polymers.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Dendrimer Nanoparticles

Dendrimers are globular macromolecules (5–10 nm) with well-defined branching architectures and surface functional groups available for further modification. The multifunctional capabilities possible through controlled synthesis methods are leading to new classes of dendrimers that can carry drug molecules, diagnostic agents, and targeting molecules. Dendrimers have remarkable molecular monodispersity and suitable pharmacokinetic properties for systemic drug delivery with cleavable chemistry for drug dissociation (C. C. Lee, MacKay, Frechet, & Szoka, 2005). Amphiphilic dendrimers are able to form micelles by self-assembly with hydrophilic groups on the surface for functionalization. Drug release kinetics are controlled through the properties of the polymer chains, which can be designed to be degraded for release of a payload. *Baker et al.* have developed "avidimers" (Majoros, Thomas, Mehta, & Baker, 2005); (Majoros, Myc, Thomas, Mehta, & Baker, 2006), which are dendrimers targeted to tumor vasculature using a methotrexatepolyamidoamine (PAMAM) bioconjugate platform functionalized with small targeting ligands (Quintana et al., 2002). Non-targeted and folate-targeted G5- PAMAM dendrimers differentially accumulated into a human KB cell line xenograft tumor model within a day (8%–10% targeted versus 2% non-targeted I.D. /g of tissues) (Kukowska-Latallo et al., 2005). Higher accumulation in the tumor resulted in the inhibition of tumor growth, lower toxicity, and longer survival time compared to free drug at equal dosage. More importantly, recent efficacy studies using targeted transferrin-cyclodextrin-siRNA nanoparticles (CALAA-01, ~70 nm) in animal models of human epithelial cancer showed tumor size reduction and differential distribution in tumors (Bartlett, Su, Hildebrandt, Weber, & Davis, 2007); (Davis, 2009); (Davis & Brewster, 2004). The preclinical data motivated further development of CALAA-01. The toxicological results reported in April 2007 for CALAA-01, which was the first targeted, polymeric nanoparticle platform in non-human primates (Heidel et al., 2007), led to the submission of an investigational new drug application and human clinical trials for solid tumor therapy in May 2008.

Polymersome Nanoparticles

Polymersomes have a structure similar to liposomes, but are composed of synthetic polymer/polypeptide amphiphiles and self-assemble to form polymer shell vesicles (~100 nm) when hydrated and extruded. Discher et al (Discher et al., 1999)described vesicles made of amphiphilic diblock copolymers with low water permeability. The hydrophilicity/hydrophobicity ratio is used to control the morphology of the nanoparticle, which can range from spherical to cylindrical. The membrane core thickness can be controlled by the molecular weight of the diblock copolymer. Polymersomes show higher stability and lateral fluidity than liposomes and the release is triggered by the degradation of the polymer chain and destabilization of the shell layer. Incubation of polymersomes in the blood showed adherence and uptake by white blood cells within 10 h. In vivo results using a breast cancer tumor xenograft model showed therapeutic efficacy after a single Intravenous injection using polymersomes loaded with paclitaxel and doxorubicin at the maximum tolerated dose (2.5 mg/kg for each drug). The tumor size was reduced within five days post injection in contrast to the free drug formulations (Ahmed et al., 2006).

Protein Nanoparticles

Protein-based drug delivery systems have recently made a big impact with albumin bound drug nanoparticles (~130 nm). The recent approval of albumin-bound paclitaxel (Abraxane, ABI-008, January 2005)

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

by the Food and Drug Administration (FDA) for metastatic breast cancer therapy, as well as multiple clinical trials currently in progress for other types of cancer, has now opened the possibility of using protein-based nanoparticles for delivery of therapeutic agents (Gradishar, 2006). Given the limiting pharmacokinetic properties and numerous side effects of Taxol (hypersensitivity), the albumin-bound paclitaxel allows the formulation of the hydrophobic drug in a solvent-free solution. Albumin is a natural noncovalent physiological transporter of molecules across endothelial barriers through a transcytosis-mediated mechanism (caveolae vesicle). Preclinical studies have shown that the concentration of paclitaxel bound to albumin in endothelial cells and in the extravascular space was significantly increased (3–10 fold) (Desai et al., 2006); (Nyman et al., 2005). Data suggests that albumin may have intrinsic targeting abilities to tumors, although the enhanced permeability and retention (EPR) effect may play an additional role in tumor accumulation. Overall, the albumin-bound paclitaxel formulation allowed higher dosages than the Taxol formulation (260 mg/m^2 vs. 175 mg/m^2, respectively) and demonstrated improved efficacy and safety (Nyman, et al., 2005). Abraxane is currently being tested as a first-line therapy or in combination with other drugs (rapamycin, verinostat, etc.) for metastatic breast cancer and other cancers that have been shown to be sensitive to taxane drugs, such as ovarian and prostate. In addition, albumin is now being tested as a platform for delivery of other molecules that have reduced water solubility, such as rapamycin (~2.5 mg/ ml). Albumin-bound rapamycin (ABI-009) has been in a clinical phase trial for the treatment of non-hematologic malignancies since January 2008.

Biological Nanoparticles

Biological nanoparticles such as bacteria are unicellular microorganisms with different shapes and sizes that encapsulate essential components of the cytoplasm as well as hydrophobic and hydrophilic molecules. One example of biological nanoparticles being evaluated for cancer therapy is a drug delivery system developed by EnGeneIC Pty Ltd called a "nanocell", which consists of anucleate globular bacteria (~400 nm). The absence of DNA prevents endogenous mutations and replication originally reported in 1967 (Adler, Fisher, Cohen, & Hardigree, 1967). It has been demonstrated that a nanocell can be efficiently loaded with molecules of different solubility and charge, such as doxorubicin, paclitaxel, and siRNA, through drug diffusion into the bacteria within a few hours (MacDiarmid et al., 2007). No signs of toxicity have been reported in large animals such as pigs and monkeys with repeated dosages at high titers, although there is the potential for an immunological response to the carrier due to the presence of lipopolysaccharide (LPS).

Inorganic Nanoparticles

Inorganic nanoparticles are primarily metal-based and have the potential to be synthesized with near monodispersity. Inorganic materials have been extensively studied for imaging using magnetic resonance and high-resolution superconducting quantum interference devices while their intrinsic properties have been explored for therapy. Several types of metal nanoparticles (Cheng et al., 2008); (Paciotti et al., 2004); (Visaria et al., 2007) are able to convert energy into heat at levels up to 70^{0}C through near-infrared light excitation or oscillating magnetic field stimulation (Johannsen et al., 2005). Iron oxide nanoparticles coated with aminosilane (Nanotherm M01) are in clinical phase II trials in Germany for brain cancer therapy and recurrent prostate cancer therapy using hyperthermia as well as thermo ablation methods. The phase I results showed that prostate tumor cells can be locally killed by

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

magnetic iron oxide nanoparticles (Johannsen, et al., 2005). Nanoparticles were injected locally using ultrasound to guide tumor injections and patients were treated once a week for 1 h over two months. The small nanoparticles (~20 nm) are able to penetrate tumors, enter cancer cells, and generate heat under magnetic fields (50 and 100 kHz), allowing treatment width between 20 and 30 cm and within a circular area of 20 cm of diameter. The authors report no dose-limiting toxicities and mild discomfort from internal heating. Similarly, silica nanoparticles coated with gold that absorb near-infrared laser energy and convert it into heat to kill solid tumors are currently under investigation in a pilot study for head and neck cancer therapy. In vivo results (Hirsch et al., 2003) of nanoshell-mediated NIR (near infrared) thermal therapy using human breast cancer xenograft models showed that the nanoparticles induced irreversible cancer tissue damage at a temperature ~40^{0}C. However, the temperature variance between different mice treated was quite significant (28–60^{0}C) and was suggested to be due to differential distribution of nanoshells in the treated volume of the tumor. In addition, the maximum recorded temperature was only ~1 mm under the skin. Recently, the same nanoparticles (150 nm) were used for brain cancer treatment in an orthotopic canine model (Schwartz et al., 2009). Tumors were killed using percutaneous infiltrated NIR fibers reaching a temperature of ~70^{0}C in tumor tissues and ~50^{0}C in normal white and grey matter, which is expected to significantly damage non-diseased areas of the brain. Surface properties and functionalities of gold nanoparticles have also been used for the delivery of surface-bound therapeutics. Aurimune (CYT-6091) is an example of tumor necrosis factor (TNF)-alpha bound to PEG-coated gold nanoparticles (~27 nm) developed by CytImmune Sciences, Inc. for solid tumor therapy (Paciotti, et al., 2004). TNF-alpha is a potent cytokine with antitumor cytotoxicity which requires incorporation into a nanocarrier formulation to reduce systemic toxicity. The results show that nanoparticle formulations delayed the tumor growth with local heating (42^{0}C for 1 h) using a SCK mammary tumor xenograft mouse model. However, the combined treatment showed a higher efficacy and suppression of intratumor blood flow (Visaria, et al., 2007). Preliminary SEM micrographs of nanoparticles accumulated in breast tumor tissue sections in contrast to healthy tissues showed possible targeting of the nanoparticles by the EPR effect. Many other formulations are still in the discovery stage using combinations of drugs such as TNF with paclitaxel, doxorubicin or interleukin-12. However, the load of therapeutic agent is reported to be several hundreds of molecules due to the surface adsorption density, which may limit the effect of the therapeutic agent. Recently, Adair's group (Kester et al., 2008); (Morgan et al., 2008) has reported the encapsulation of organic molecules in calcium phosphate nanocomposite particles (~27 nm) for intracellular imaging and delivery. Calcium phosphate-based nanoparticles are biocompatible and their pH dissolution properties can be used for controlled release of molecules in the acidic tumor environment. In vitro studies show high uptake of the nanoparticles in bovine aortic endothelial cells and the delivery of hexanoyl-ceramide (Cer-6) to human vascular smooth muscle cells showed 100% inhibition of cell growth at 200 nM of drug (Kester, et al., 2008). This technology is now being developed by Keystone Nano for imaging and delivery of therapeutic agents. Non-specific accumulation into healthy tissues is always a concern for nanoparticle drug delivery systems. Using local sensitization through light or temperature may reduce overall toxicity, but it is expected to damage adjacent healthy tissues as well. Ultimately, inorganic particles may not provide advantages over other types of nanoparticles for systemic targeting of cancer cells because they are not biodegradable, have low payloads, and have no controlled release properties.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Hybrid Nanoparticles

Hybrid nanoparticles are recently developed nanocarriers that combine advantages from existing systems with well-characterized properties to form lipid–polymer nanoparticles and solid liposomal nanoparticles. Hybrid nanoparticles are composed of at least two different materials to form the core and the corona structure. In general, metallic and polymeric materials form the core and are coated with a single or multiple lipid layers to form a protecting membrane (corona) similar to a liposome or micelle. Recently Farokhzad's group (L. Zhang et al., 2008) have developed hybrid nanoparticles for cancer therapy. Sasisekharan and co-workers (Sengupta et al., 2005) have reported PLGA-core nanoparticles coated with a bi-phospholipid layer to carry multiple drugs for cancer therapy using melanoma and Lewis lung carcinoma models. In their system, doxorubicin is conjugated to PLGA to form the core of the nanoparticle (~1% load by weight of doxorubicin, 70% encapsulation efficiency) while an anti-angiogenesis drug, combrestatin, is mixed with phospholipids and encapsulated in the lipid bi-layer during the self-assembly process to form nanoparticles (~200 nm) described as "nanocells". The drugs were release at different rates over a period of ~3 days, with combrestatin released first to reduce vascular density in the tumor followed by the release of doxorubicin to kill the cancer cells. The results showed a significant delay in tumor growth and increased survival time in cancer models, suggesting accumulation of the nanocell by the EPR effect and added therapeutic value by delivering multiple drugs. The nanocell technology is now in preclinical development by Cerulean Pharma. Others have reported solid-lipid nanoparticles using different polymers and formulations in vitro and in vivo for combination therapy. Recently, (Thevenot, Troutier, David, Delair, & Ladaviere, 2007) described a mechanism for the encapsulation of a hydrophobic polymer core (PLA) in PEG-liposomes. As part of the work, the importance of PEG chain length to sterically stabilize lipoparticles with optimal colloidal stability was demonstrated (PEG (5 kDa) at 10% of lipid content).

FUTURE PERSPECTIVE ON PASSIVE TARGETING

The last 30 years have shown that EPR plays an important role in the delivery of macromolecules and nanomaterials to tumors. Nevertheless, a full understanding of the degree of interpatient variability and the importance of tumor heterogeneity in the EPR effect in humans has yet to be established (Prabhakar et al., 2013). As we unravel which parameters are the most important to improve the distribution of chemotherapeutics to tumor cells, the individual patients and the cancer types benefiting the most from carrier-mediated drug delivery might be identified. Furthermore, it is now appreciated that the physicochemical properties of the nanomaterial are as important as the tumor biology. Meaningful exploitation of the EPR effect in humans will require precise understanding and control of the physicochemical properties of nanomaterials (L. Zhang, et al., 2008) and, possibly artificial fine tuning of the tumor microenvironment (Carmeliet & Jain, 2011). In the era of modern cancer biology, assessing the tumor microenvironment in individual patients and predicting their susceptibility to the EPR effect may eventually become the mainstay of therapy when choosing between therapeutic regimens. Although it is hard to currently predict which markers better correlate with NPs efficacy in humans, relationships will most likely delineate as tumor genotyping and bioassays continue to progress. Ideally, a broad biomarker correlating the susceptibility of tumors to nanomedicines in general (i.e., levels of tumor-associated macrophages; levels of ECM proteins/GAG; tumor infiltration of specific proteins/cells; endothelial

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

expression of surface biomarkers) would certainly facilitate the screening of patients to optimize therapy.In parallel, the EPR phenomenon affects the distribution of the drug-carrier to the tumor without necessarily increasing the ability of the drug to reach its pharmacological target. The optimization of NPs' efficacy therefore also involves optimal drug release rates whether through controlled diffusional release (X. Q. Zhang et al., 2012), covalent conjugation of the drug to a polymer backbone (Kolishetti et al., 2010); (Aryal, Hu, & Zhang, 2010) or environment-triggered release(Bertrand, Fleischer, Wasan, & Leroux, 2009); (Gao, Langer, & Farokhzad, 2010). Without specific affinity of the nanomaterial for the cancer cells, the chemotherapeutic payloads will have to reach their pharmacological targets by their own means or risk diffusing back into the vasculature (Dreher et al., 2006). In some instances, drugs exhibit sufficient affinity for their pharmacological target to remain trapped in the tumor for prolonged periods of time. For example, the high affinity of docetaxel for the microtubules translates into very low flux from the tumor; consequently, the elimination half-life of the drug from tumors is approximately 15-20 times higher than its elimination from the blood and normal tissue (Bissery, Nohynek, Sanderink, & Lavelle, 1995). The behaviour of drugs and their affinity for the intratumoral environment need to be taken into account individually when designing passively-targeted NPs and the optimal drug release profiles should be optimized on a case-by-case basis. In this regard, tools that can assess the interactions between drugs and their substrates in tumors will be of particular interest (Ullal et al., 2011). Similarly, most macromolecular drugs (including nucleic acids and some proteins) cannot readily permeate through the cell membrane and reach their pharmacological target. For these classes of API the modification of the NPs with targeting ligands may be more appropriate. Active targeting increases the affinity of the NPs for tumor cells, increasing its tumor residence times and allowing the drug-loaded NPs to efficiently enter the cells through receptor mediated endocytosis. The principles and future development of active targeting using affinity ligands on the surface of NPs will be presented in the next section.

IMMEDIATE FUTURE OF CANCER NANOMEDICINES

More than 30 years after the first observation of the EPR effect, we are just unravelling the mechanisms behind the distribution of macro- molecules and nanomaterials inside the tumor. At the preclinical level, further investigation on how physicochemical properties affect the tumor deposition and intratumoral diffusion of nanomaterials might continue to influence the design of future cancer therapeutic NPs. The evolution of chemistry and material sciences constantly pushes back the limits of nanomaterial synthesis. As the tumor distribution of materials with different sizes, shapes, charges and physicochemical properties continue to be evaluated, the body of knowledge on the EPR effect can expand significantly. The key to improve our understanding lies in the adequate design of studies with respect to the current animal models and their limitations. The clinical development of future anticancer NPs also requires strict toxicological validation; it is therefore probable that biodegradable and biocompatible materials might continue to play a major role in the future development of cancer therapeutic NPs. Thanks to the experience acquired on already approved nanomedicines, it is now appreciated that the susceptibility of patients to nano sized drug carriers not only depends on the type of cancer, but also on individual tumor biology aspects. Until now, the most successful treatments resulted from increased tolerability to the NPs (e.g., the maximum tolerated dose of Nab-PTX is 1.5 that of solvent-formulated PTX). For PEGylated liposomal DOX, the major survival benefit was witnessed as a second-line treatment for recurring ovarian cancer where the altered toxicity profile of the drug allows repeated administrations over multiple

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

cycles. In a clinical setting, relying on the EPR effect alone to significantly impact the efficacy of cancer nanomedicines in all patients seems increasingly unrealistic; strategies must therefore be investigated to improve the tumor distribution of the drugs and their efficacy. To that end, screening patients a priori for their susceptibility to the EPR effect and preserving nanomedicines for those who would benefit the most from them might seem a viable, short term approach. It is also important to fully understand what dictates this increased predisposition to the EPR effect in humans. Ideally, clinical studies would evaluate nanomedicines in combination with biomarker screening and patient phenotyping to assess which characteristics are good predictors of efficacy. For this purpose, the clinical imaging community might provide useful, minimally-invasive tools. The ability to image in real-time what is happening in the tumor will probably prove very valuable in studying the tumor distribution of nanomaterials. Clinical initiatives trying to correlate tumor uptake of imaging nanomaterials with clinical efficacy of therapeutic NPs might answer many persisting questions. In parallel, recent advances suggest that tumor remodelling can be achieved with commercially -available drugs (e.g., antiangiogenic bevacizumab or antifibrotic losartan). Therefore, minimal additional preclinical work could provide the bases for a pilot study to compare the combination of such treatment with cytotoxic nanocarriers in patients. If successful, these studies might provide viable clinical regimens to better fight tumors with the current available tools. Active targeting is another tool that could modify the paradigm of our fight against cancer using nanotechnologies. Given the difficulties of immediately ascertaining the impact of active targeting in humans, comprehensive work to really highlight the advantages of using NPs decorated with surface ligands might need to be conducted at the pre-clinical stage. To that end, the development of animal models which offer good predictive values for what is observed in the clinics is paramount and the standards of preclinical research might have to shift toward the use of orthotopic or genetically engineered models. Such models might help understand the impact of each physicochemical property on the cellular uptake and efficacy of the NPs. Since the tumor distribution of actively-targeted NPs seems strongly associated with the EPR effect, it will be interesting to see how insights obtained by studying both passive and active targeting strategies might converge to lead to more efficacious treatments. Active targeting strategies might possibly show added benefits as a means for active ingredients to cross physiological barriers, whether they are as simple as cell membranes, or as complex as the gastrointestinal epithelia or the blood brain barrier. In parallel, as tumor biology progresses to expose new molecular targets on cancer cells, improvements in ligand screening techniques should facilitate the isolation of highly specific ligands with strong affinities. While antibodies and proteins will probably continue to be beneficial to establish proof-of-concepts because of their availability, efficiency and applicability, peptides and small molecules might play an increasing role in the development of more advanced NPs. The next few years will hopefully demonstrate the clinical validation of actively-targeted NPs. Here again, in the context of personalized medicine, the challenge will be to clearly identify the patients who might benefit from actively-targeted treatments. When assessing the efficacy of a novel treatment in a cohort, it is crucial to make sure that receptive populations are not concealed by more abundant non-responsive phenotypes. The risks of having the patients' characteristics impact response rates increases with the complexity of the given therapy. When treatments highly depend on specific metabolic pathways or the presence of a surface receptor to be efficacious, mutations and changes in phenotypes must be closely monitored to ascertain that optimal characteristics are maintained throughout the treatment. In this regard, advances in genotyping and phenotyping techniques should prove very insightful toward novel blueprints for more advanced clinical trials. Hence, there is good hope that the NPs currently under study might reveal very

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

useful tools to fight cancer. Such breakthroughs might contribute to the blooming of cancer nanomedicines in general and actively-targeted platforms, in particular.

REFERENCES

Adler, H. I., Fisher, W. D., Cohen, A., & Hardigree, A. A. (1967). Miniature escherichia coli cells deficient in DNA. *Proceedings of the National Academy of Sciences of the United States of America*, *57*(2), 321–326. doi:10.1073/pnas.57.2.321 PMID:16591472

Ahmed, F., Pakunlu, R. I., Srinivas, G., Brannan, A., Bates, F., & Klein, M. L. et al. (2006). Shrinkage of a rapidly growing tumor by drug-loaded polymersomes: pH-triggered release through copolymer degradation. *Molecular Pharmaceutics*, *3*(3), 340–350. doi:10.1021/mp050103u PMID:16749866

Alexis, F., Basto, P., Levy-Nissenbaum, E., Radovic-Moreno, A. F., Zhang, L., & Pridgen, E. et al. (2008). HER-2-targeted nanoparticle-affibody bioconjugates for cancer therapy. *ChemMedChem*, *3*(12), 1839–1843. doi:10.1002/cmdc.200800122 PMID:19012296

Alexis, F., Pridgen, E., Molnar, L. K., & Farokhzad, O. C. (2008). Factors affecting the clearance and biodistribution of polymeric nanoparticles. *Molecular Pharmaceutics*, *5*(4), 505–515. doi:10.1021/mp800051m PMID:18672949

Aliabadi, H. M., Shahin, M., Brocks, D. R., & Lavasanifar, A. (2008). Disposition of drugs in block copolymer micelle delivery systems: from discovery to recovery. *Clin Pharmacokinet, 47*(10), 619-634. doi: 10.2165/00003088-200847100-00001

Andresen, T. L., Davidsen, J., Begtrup, M., Mouritsen, O. G., & Jorgensen, K. (2004). Enzymatic release of antitumor ether lipids by specific phospholipase A2 activation of liposome-forming prodrugs. *Journal of Medicinal Chemistry*, *47*(7), 1694–1703. doi:10.1021/jm031029r PMID:15027860

Aryal, S., Hu, C. M., & Zhang, L. (2010). Combinatorial drug conjugation enables nanoparticle dual-drug delivery. *Small*, *6*(13), 1442–1448. doi:10.1002/smll.201000631 PMID:20564488

Bartlett, D. W., Su, H., Hildebrandt, I. J., Weber, W. A., & Davis, M. E. (2007). Impact of tumor-specific targeting on the biodistribution and efficacy of siRNA nanoparticles measured by multimodality in vivo imaging. *Proc Natl Acad Sci U S A, 104*(39), 15549-15554. doi: 10.1073/pnas.0707461104

Bates, D. O., Hillman, N. J., Williams, B., Neal, C. R., & Pocock, T. M. (2002). Regulation of microvascular permeability by vascular endothelial growth factors. *Journal of Anatomy*, *200*(6), 581–597. doi:10.1046/j.1469-7580.2002.00066.x PMID:12162726

Bertrand, N., Fleischer, J. G., Wasan, K. M., & Leroux, J. C. (2009). Pharmacokinetics and biodistribution of N-isopropylacrylamide copolymers for the design of pH-sensitive liposomes. *Biomaterials, 30*(13), 2598-2605. doi: 10.1016/j.biomaterials.2008.12.082

Bertrand, N., & Leroux, J. C. (2012). The journey of a drug-carrier in the body: an anatomo-physiological perspective. *J Control Release, 161*(2), 152-163. doi: 10.1016/j.jconrel.2011.09.098

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Bierie, B., & Moses, H. L. (2006). Tumour microenvironment: TGFbeta: the molecular Jekyll and Hyde of cancer. *Nat Rev Cancer, 6*(7), 506-520. doi: 10.1038/nrc1926

Bissell, M. J., & Labarge, M. A. (2005). Context, tissue plasticity, and cancer: are tumor stem cells also regulated by the microenvironment? *Cancer Cell, 7*(1), 17-23. doi: 10.1016/j.ccr.2004.12.013

Bissery, M. C., Nohynek, G., Sanderink, G. J., & Lavelle, F. (1995). Docetaxel (Taxotere): A review of preclinical and clinical experience. Part I: Preclinical experience. *Anticancer Drugs, 6*(3), 339-355, 363-338.

Boddy, A. V., Plummer, E. R., Todd, R., Sludden, J., Griffin, M., Robson, L., et al. (2005). A phase I and pharmacokinetic study of paclitaxel poliglumex (XYOTAX), investigating both 3-weekly and 2-weekly schedules. *Clin Cancer Res, 11*(21), 7834-7840. doi: 10.1158/1078-0432.CCR-05-0803

Brigger, I., Dubernet, C., & Couvreur, P. (2002). Nanoparticles in cancer therapy and diagnosis. *Adv Drug Deliv Rev, 54*(5), 631-651.

Cairns, R., Papandreou, I., & Denko, N. (2006). Overcoming physiologic barriers to cancer treatment by molecularly targeting the tumor microenvironment. *Mol Cancer Res, 4*(2), 61-70. doi: 10.1158/1541-7786.MCR-06-0002

Campone, M., Rademaker-Lakhai, J. M., Bennouna, J., Howell, S. B., Nowotnik, D. P., Beijnen, J. H., & Schellens, J. H. M. (2007). Phase I and pharmacokinetic trial of AP5346, a DACH-platinum-polymer conjugate, administered weekly for three out of every 4 weeks to advanced solid tumor patients. *Cancer Chemotherapy and Pharmacology*, *60*(4), 523–533. doi:10.1007/s00280-006-0397-0 PMID:17308894

Carmeliet, P., & Jain, R. K. (2011). Principles and mechanisms of vessel normalization for cancer and other angiogenic diseases. *Nat Rev Drug Discov, 10*(6), 417-427. doi: 10.1038/nrd3455

Castor, T. P. (2005). Phospholipid nanosomes. *Current Drug Delivery*, *2*(4), 329–340. doi:10.2174/156720105774370195 PMID:16305436

Chauhan, V. P., Popovic, Z., Chen, O., Cui, J., Fukumura, D., Bawendi, M. G., & Jain, R. K. (2011). Fluorescent nanorods and nanospheres for real-time in vivo probing of nanoparticle shape-dependent tumor penetration. *Angewandte Chemie (International ed. in English)*, *50*(48), 11417–11420. doi:10.1002/anie.201104449 PMID:22113800

Cheng, Y., C. Samia, A., Meyers, J. D., Panagopoulos, I., Fei, B., & Burda, C. (2008, August). Highly Efficient Drug Delivery with Gold Nanoparticle Vectors for *in Vivo* Photodynamic Therapy of Cancer. *Journal of the American Chemical Society*, *130*(32), 10643–10647. doi:10.1021/ja801631c PMID:18642918

Couvreur, P., Kante, B., Roland, M., & Speiser, P. (1979). Adsorption of antineoplastic drugs to polyalkylcyanoacrylate nanoparticles and their release in calf serum. *Journal of Pharmaceutical Sciences*, *68*(12), 1521–1524. doi:10.1002/jps.2600681215 PMID:529043

Cullis, P. R., & Chonn, A. (1998). Recent advances in liposome technologies and their applications for systemic gene delivery. *Adv Drug Deliv Rev, 30*(1-3), 73-83. doi:

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Davis, M. E. (2009). The first targeted delivery of siRNA in humans via a self-assembling, cyclodextrin polymer-based nanoparticle: From concept to clinic. *Molecular Pharmaceutics*, *6*(3), 659–668. doi:10.1021/mp900015y PMID:19267452

Davis, M. E., & Brewster, M. E. (2004). Cyclodextrin-based pharmaceutics: past, present and future. *Nat Rev Drug Discov, 3*(12), 1023-1035. doi: 10.1038/nrd1576

Davis, M. E., Chen, Z. G., & Shin, D. M. (2008). Nanoparticle therapeutics: an emerging treatment modality for cancer. *Nat Rev Drug Discov, 7*(9), 771-782. doi: 10.1038/nrd2614

Davis, M. E., Zuckerman, J. E., Choi, C. H., Seligson, D., Tolcher, A., Alabi, C. A., et al. (2010). Evidence of RNAi in humans from systemically administered siRNA via targeted nanoparticles. *Nature, 464*(7291), 1067-1070. doi: 10.1038/nature08956

des Rieux, A., Pourcelle, V., Cani, P. D., Marchand-Brynaert, J., & Preat, V. (2013). Targeted nanoparticles with novel non-peptidic ligands for oral delivery. *Adv Drug Deliv Rev, 65*(6), 833-844. doi: 10.1016/j.addr.2013.01.002

Desai, N., Trieu, V., Yao, Z., Louie, L., Ci, S., Yang, A., et al. (2006). Increased antitumor activity, intratumor paclitaxel concentrations, and endothelial cell transport of cremophor-free, albumin-bound paclitaxel, ABI-007, compared with cremophor-based paclitaxel. *Clin Cancer Res, 12*(4), 1317-1324. doi: 10.1158/1078-0432.CCR-05-1634

Discher, B. M., Won, Y. Y., Ege, D. S., Lee, J. C., Bates, F. S., & Discher, D. E. et al. (1999). Polymersomes: Tough vesicles made from diblock copolymers. *Science*, *284*(5417), 1143–1146. doi:10.1126/science.284.5417.1143 PMID:10325219

Dreher, M. R., Liu, W., Michelich, C. R., Dewhirst, M. W., Yuan, F., & Chilkoti, A. (2006). Tumor vascular permeability, accumulation, and penetration of macromolecular drug carriers. *J Natl Cancer Inst, 98*(5), 335-344. doi: 10.1093/jnci/djj070

Euliss, L. E., DuPont, J. A., Gratton, S., & DeSimone, J. (2006). Imparting size, shape, and composition control of materials for nanomedicine. *Chemical Society Reviews*, *35*(11), 1095–1104. doi:10.1039/b600913c PMID:17057838

Farokhzad, O. C. (2008). Nanotechnology for drug delivery: The perfect partnership. *Expert Opinion on Drug Delivery*, *5*(9), 927–929. doi:10.1517/17425247.5.9.927 PMID:18754745

Farokhzad, O. C., & Langer, R. (2006). Nanomedicine: developing smarter therapeutic and diagnostic modalities. *Adv Drug Deliv Rev, 58*(14), 1456-1459. doi: 10.1016/j.addr.2006.09.011

Ferrari, M. (2005). Cancer nanotechnology: opportunities and challenges. *Nat Rev Cancer, 5*(3), 161-171. doi: 10.1038/nrc1566

Florence, A. T. (2012). "Targeting" nanoparticles: the constraints of physical laws and physical barriers. *J Control Release, 164*(2), 115-124. doi: 10.1016/j.jconrel.2012.03.022

Fox, J. L. (2000). Researchers discuss NIH's nanotechnology initiative. *Nature Biotechnology*, *18*(8), 821. doi:10.1038/78418 PMID:10932144

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Freitas, R. A., Jr. (2005). What is nanomedicine? *Nanomedicine, 1*(1), 2-9. doi: 10.1016/j.nano.2004.11.003

Gao, W., Langer, R., & Farokhzad, O. C. (2010). Poly(ethylene glycol) with observable shedding. *Angewandte Chemie (International ed. in English)*, *49*(37), 6567–6571. doi:10.1002/anie.201001868 PMID:20677297

Geng, Y., Dalhaimer, P., Cai, S., Tsai, R., Tewari, M., Minko, T., et al. (2007). Shape effects of filaments versus spherical particles in flow and drug delivery. *Nat Nanotechnol, 2*(4), 249-255. doi: 10.1038/nnano.2007.70

Gradishar, W. J. (2006). Albumin-bound paclitaxel: A next-generation taxane. *Expert Opinion on Pharmacotherapy*, *7*(8), 1041–1053. doi:10.1517/14656566.7.8.1041 PMID:16722814

Gratton, S. E., Napier, M. E., Ropp, P. A., Tian, S., & DeSimone, J. M. (2008). Microfabricated particles for engineered drug therapies: Elucidation into the mechanisms of cellular internalization of PRINT particles. *Pharmaceutical Research*, *25*(12), 2845–2852. doi:10.1007/s11095-008-9654-8 PMID:18592353

Gratton, S. E., Pohlhaus, P. D., Lee, J., Guo, J., Cho, M. J., & Desimone, J. M. (2007). Nanofabricated particles for engineered drug therapies: a preliminary biodistribution study of PRINT nanoparticles. *J Control Release, 121*(1-2), 10-18. doi: 10.1016/j.jconrel.2007.05.027

Gref, R., Luck, M., Quellec, P., Marchand, M., Dellacherie, E., Harnisch, S., et al. (2000). 'Stealth' corona-core nanoparticles surface modified by polyethylene glycol (PEG): Influences of the corona (PEG chain length and surface density) and of the core composition on phagocytic uptake and plasma protein adsorption. *Colloids Surf B Biointerfaces, 18*(3-4), 301-313.

Gu, F., Zhang, L., Teply, B. A., Mann, N., Wang, A., Radovic-Moreno, A. F., et al. (2008). Precise engineering of targeted nanoparticles by using self-assembled biointegrated block copolymers. *Proc Natl Acad Sci U S A, 105*(7), 2586-2591. doi: 10.1073/pnas.0711714105

Guo, X., Andrew MacKay, J., & Szoka, F. C. Jr. (2003). Mechanism of pH-triggered collapse of phosphatidylethanolamine liposomes stabilized by an ortho ester polyethyleneglycol lipid. *Biophysical Journal*, *84*(3), 1784–1795. doi:10.1016/S0006-3495(03)74986-8 PMID:12609880

Hamaguchi, T., Kato, K., Yasui, H., Morizane, C., Ikeda, M., Ueno, H., et al. (2007). A phase I and pharmacokinetic study of NK105, a paclitaxel-incorporating micellar nanoparticle formulation. *Br J Cancer, 97*(2), 170-176. doi: 10.1038/sj.bjc.6603855

Hashizume, H., Baluk, P., Morikawa, S., McLean, J. W., Thurston, G., Roberge, S., et al. (2000). Openings between defective endothelial cells explain tumor vessel leakiness. *Am J Pathol, 156*(4), 1363-1380. doi: 10.1016/S0002-9440(10)65006-7

Heidel, J. D., Yu, Z., Liu, J. Y., Rele, S. M., Liang, Y., Zeidan, R. K., et al. (2007). Administration in non-human primates of escalating intravenous doses of targeted nanoparticles containing ribonucleotide reductase subunit M2 siRNA. *Proc Natl Acad Sci U SA, 104*(14), 5715-5721. doi: 10.1073/pnas.0701458104

Hirsch, L. R., Stafford, R. J., Bankson, J. A., Sershen, S. R., Rivera, B., Price, R. E., et al. (2003). Nanoshell-mediated near-infrared thermal therapy of tumors under magnetic resonance guidance. *Proc Natl Acad Sci U S A, 100*(23), 13549-13554. doi: 10.1073/pnas.2232479100

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Hobbs, S. K., Monsky, W. L., Yuan, F., Roberts, W. G., Griffith, L., Torchilin, V. P., & Jain, R. K. (1998). Regulation of transport pathways in tumor vessels: Role of tumor type and microenvironment. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(8), 4607–4612. doi:10.1073/pnas.95.8.4607 PMID:9539785

Homsi, J., Simon, G. R., Garrett, C. R., Springett, G., De Conti, R., Chiappori, A. A., et al. (2007). Phase I trial of poly-L-glutamate camptothecin (CT-2106) administered weekly in patients with advanced solid malignancies. *Clin Cancer Res, 13*(19), 5855-5861. doi: 10.1158/1078-0432.CCR-06-2821

Jain, K. K. (2008). Nanomedicine: application of nanobiotechnology in medical practice. *Med Princ Pract, 17*(2), 89-101. doi: 10.1159/000112961

Jain, R. K. (1998). The next frontier of molecular medicine: Delivery of therapeutics. *Nature Medicine*, *4*(6), 655–657. doi:10.1038/nm0698-655 PMID:9623964

Jain, R. K., & Stylianopoulos, T. (2010). Delivering nanomedicine to solid tumors. *Nat Rev Clin Oncol, 7*(11), 653-664. doi: 10.1038/nrclinonc.2010.139

Jemal, A., Siegel, R., Xu, J., & Ward, E. (2010). Cancer statistics, 2010. *CA Cancer J Clin, 60*(5), 277-300. doi: 10.3322/caac.20073

Jensen, S. S., Andresen, T. L., Davidsen, J., Hoyrup, P., Shnyder, S. D., Bibby, M. C., et al. (2004). Secretory phospholipase A2 as a tumor-specific trigger for targeted delivery of a novel class of liposomal prodrug anticancer etherlipids. *Mol Cancer Ther, 3*(11), 1451-1458.

Jiang, W., Kim, B. Y., Rutka, J. T., & Chan, W. C. (2007). Advances and challenges of nanotechnology-based drug delivery systems. *Expert Opinion on Drug Delivery*, *4*(6), 621–633. doi:10.1517/17425247.4.6.621 PMID:17970665

Johannsen, M., Gneveckow, U., Eckelt, L., Feussner, A., Waldofner, N., & Scholz, R. et al. (2005). Clinical hyperthermia of prostate cancer using magnetic nanoparticles: Presentation of a new interstitial technique. *International Journal of Hyperthermia*, *21*(7), 637–647. doi:10.1080/02656730500158360 PMID:16304715

Jorgensen, K., Davidsen, J., & Mouritsen, O. G. (2002). Biophysical mechanisms of phospholipase A2 activation and their use in liposome-based drug delivery. *FEBS Lett, 531*(1), 23-27.

Kamaly, N., Xiao, Z., Valencia, P. M., Radovic-Moreno, A. F., & Farokhzad, O. C. (2012). Targeted polymeric therapeutic nanoparticles: design, development and clinical translation. *Chemical Society Reviews*, *41*(7), 2971–3010. doi:10.1039/c2cs15344k PMID:22388185

Kawasaki, E. S., & Player, A. (2005). Nanotechnology, nanomedicine, and the development of new, effective therapies for cancer. *Nanomedicine, 1*(2), 101-109. doi: 10.1016/j.nano.2005.03.002

Kelly, J. Y., & DeSimone, J. M. (2008). Shape-specific, monodisperse nano-molding of protein particles. *Journal of the American Chemical Society*, *130*(16), 5438–5439. doi:10.1021/ja8014428 PMID:18376832

Kester, M., Heakal, Y., Fox, T., Sharma, A., Robertson, G. P., & Morgan, T. T. et al. (2008). Calcium phosphate nanocomposite particles for in vitro imaging and encapsulated chemotherapeutic drug delivery to cancer cells. *Nano Letters*, *8*(12), 4116–4121. doi:10.1021/nl802098g PMID:19367878

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Kim, D. W., Kim, S. Y., Kim, H. K., Kim, S. W., Shin, S. W., Kim, J. S., et al. (2007). Multicenter phase II trial of Genexol-PM, a novel Cremophor-free, polymeric micelle formulation of paclitaxel, with cisplatin in patients with advanced non-small-cell lung cancer. *Ann Oncol, 18*(12), 2009-2014. doi: 10.1093/annonc/mdm374

Kim, K., Lee, M., Park, H., Kim, J. H., Kim, S., & Chung, H. et al. (2006). Cell-permeable and biocompatible polymeric nanoparticles for apoptosis imaging. *Journal of the American Chemical Society*, *128*(11), 3490–3491. doi:10.1021/ja057712f PMID:16536501

Kim, S. C., Kim, D. W., Shim, Y. H., Bang, J. S., Oh, H. S., Wan Kim, S., et al. (2001). In vivo evaluation of polymeric micellar paclitaxel formulation: toxicity and efficacy. *J Control Release, 72*(1-3), 191-202.

Kim, T. Y., Kim, D. W., Chung, J. Y., Shin, S. G., Kim, S. C., Heo, D. S., et al. (2004). Phase I and pharmacokinetic study of Genexol-PM, a cremophor-free, polymeric micelle-formulated paclitaxel, in patients with advanced malignancies. *Clin Cancer Res, 10*(11), 3708-3716. doi: 10.1158/1078-0432.CCR-03-0655

Kocer, A. (2007). A remote controlled valve in liposomes for triggered liposomal release. *J Liposome Res, 17*(3-4), 219-225. doi: 10.1080/08982100701528203

Kolishetti, N., Dhar, S., Valencia, P. M., Lin, L. Q., Karnik, R., Lippard, S. J., et al. (2010). Engineering of self-assembled nanoparticle platform for precisely controlled combination drug therapy. *Proc Natl Acad Sci U S A, 107*(42), 17939-17944. doi: 10.1073/pnas.1011368107

Koshkaryev, A., Sawant, R., Deshpande, M., & Torchilin, V. (2013). Immunoconjugates and long circulating systems: origins, current state of the art and future directions. *Adv Drug Deliv Rev, 65*(1), 24-35. doi: 10.1016/j.addr.2012.08.009

Kukowska-Latallo, J. F., Candido, K. A., Cao, Z., Nigavekar, S. S., Majoros, I. J., Thomas, T. P., et al. (2005). Nanoparticle targeting of anticancer drug improves therapeutic response in animal model of human epithelial cancer. *Cancer Res, 65*(12), 5317-5324. doi: 10.1158/0008-5472.CAN-04-3921

Langer, R. (1998). Drug delivery and targeting. *Nature*, *392*(6679Suppl), 5–10. PMID:9579855

Langer, R., & Folkman, J. (1976). Polymers for the sustained release of proteins and other macromolecules. *Nature*, *263*(5580), 797–800. doi:10.1038/263797a0 PMID:995197

Lanza, G., Winter, P., Cyrus, T., Caruthers, S., Marsh, J., Hughes, M., et al. (2006). Nanomedicine opportunities in cardiology. *Ann N Y Acad Sci, 1080*, 451-465. doi: 10.1196/annals.1380.034

Lee, C. C., MacKay, J. A., Frechet, J. M., & Szoka, F. C. (2005). Designing dendrimers for biological applications. *Nat Biotechnol, 23*(12), 1517-1526. doi: 10.1038/nbt1171

Lee, K. S., Chung, H. C., Im, S. A., Park, Y. H., Kim, C. S., & Kim, S. B. et al. (2008). Multicenter phase II trial of Genexol-PM, a Cremophor-free, polymeric micelle formulation of paclitaxel, in patients with metastatic breast cancer. *Breast Cancer Research and Treatment*, *108*(2), 241–250. doi:10.1007/s10549-007-9591-y PMID:17476588

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Levy-Nissenbaum, E., Radovic-Moreno, A. F., Wang, A. Z., Langer, R., & Farokhzad, O. C. (2008). Nanotechnology and aptamers: applications in drug delivery. *Trends Biotechnol, 26*(8), 442-449. doi: 10.1016/j.tibtech.2008.04.006

Liggins, R. T., & Burt, H. M. (2002). Polyether-polyester diblock copolymers for the preparation of paclitaxel loaded polymeric micelle formulations. *Adv Drug Deliv Rev, 54*(2), 191-202.

MacDiarmid, J. A., Mugridge, N. B., Weiss, J. C., Phillips, L., Burn, A. L., Paulin, R. P., et al. (2007). Bacterially derived 400 nm particles for encapsulation and cancer cell targeting of chemotherapeutics. *Cancer Cell, 11*(5), 431-445. doi: 10.1016/j.ccr.2007.03.012

Maeda, H. (2001). SMANCS and polymer-conjugated macromolecular drugs: advantages in cancer chemotherapy. *Adv Drug Deliv Rev, 46*(1-3), 169-185.

Majoros, I. J., Myc, A., Thomas, T., Mehta, C. B., & Baker, J. R. Jr. (2006). PAMAM dendrimer-based multifunctional conjugate for cancer therapy: synthesis, characterization, and functionality. *Biomacromolecules*, *7*(2), 572–579. doi:10.1021/bm0506142 PMID:16471932

Majoros, I. J., Thomas, T. P., Mehta, C. B., & Baker, J. R. Jr. (2005). Poly(amidoamine) dendrimer-based multifunctional engineered nanodevice for cancer therapy. *Journal of Medicinal Chemistry*, *48*(19), 5892–5899. doi:10.1021/jm0401863 PMID:16161993

Matsumura, M. (1986). A new concept for macromolecular therapeutics in cancer chemotherapy: Mechanism of tumoritropic accumulation of proteins and the antitumor agent smancs. *Cancer Research*, *46*, 6387–6392. PMID:2946403

Matsumura, Y. (2008). Poly (amino acid) micelle nanocarriers in preclinical and clinical studies. *Adv Drug Deliv Rev, 60*(8), 899-914. doi: 10.1016/j.addr.2007.11.010

Matsumura, Y., Hamaguchi, T., Ura, T., Muro, K., Yamada, Y., Shimada, Y., et al. (2004). Phase I clinical trial and pharmacokinetic evaluation of NK911, a micelle-encapsulated doxorubicin. *Br J Cancer, 91*(10), 1775-1781. doi: 10.1038/sj.bjc.6602204

Moein Moghimi, S. (2006). Recent developments in polymeric nanoparticle engineering and their applications in experimental and clinical oncology. *Anti-cancer Agents in Medicinal Chemistry*, *6*(6), 553–561. doi:10.2174/187152006778699130 PMID:17100563

Moghimi, S. M., Hunter, A. C., & Murray, J. C. (2001). Long-circulating and target-specific nanoparticles: Theory to practice. *Pharmacological Reviews*, *53*(2), 283–318. PMID:11356986

Moghimi, S. M., Hunter, A. C., & Murray, J. C. (2005). Nanomedicine: current status and future prospects. *FASEB J, 19*(3), 311-330. doi: 10.1096/fj.04-2747rev

Montet, X., Weissleder, R., & Josephson, L. (2006). Imaging pancreatic cancer with a peptide-nanoparticle conjugate targeted to normal pancreas. *Bioconjugate Chemistry*, *17*(4), 905–911. doi:10.1021/bc060035+ PMID:16848396

Morgan, T. T., Muddana, H. S., Altinoglu, E. I., Rouse, S. M., Tabakovic, A., & Tabouillot, T. et al. (2008). Encapsulation of organic molecules in calcium phosphate nanocomposite particles for intracellular imaging and drug delivery. *Nano Letters*, *8*(12), 4108–4115. doi:10.1021/nl8019888 PMID:19367837

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Nasongkla, N., Bey, E., Ren, J., Ai, H., Khemtong, C., & Guthi, J. S. et al. (2006). Multifunctional polymeric micelles as cancer-targeted, MRI-ultrasensitive drug delivery systems. *Nano Letters*, *6*(11), 2427–2430. doi:10.1021/nl061412u PMID:17090068

Nyman, D. W., Campbell, K. J., Hersh, E., Long, K., Richardson, K., Trieu, V., et al. (2005). Phase I and pharmacokinetics trial of ABI-007, a novel nanoparticle formulation of paclitaxel in patients with advanced nonhematologic malignancies. *J Clin Oncol, 23*(31), 7785-7793. doi: 10.1200/JCO.2004.00.6148

Olson, R. D., & Mushlin, P. S. (1990). Doxorubicin cardiotoxicity: Analysis of prevailing hypotheses. *FASEB journal: official publication of the Federation of American Societies for Experimental Biology*, *4*(13), 3076–3086. PMID:2210154

Paciotti, G. F., Myer, L., Weinreich, D., Goia, D., Pavel, N., McLaughlin, R. E., et al. (2004). Colloidal gold: a novel nanoparticle vector for tumor directed drug delivery. *Drug Deliv, 11*(3), 169-183. doi: 10.1080/10717540490433895

Padera, T. P., Stoll, B. R., Tooredman, J. B., Capen, D., di Tomaso, E., & Jain, R. K. (2004). Pathology: cancer cells compress intratumour vessels. *Nature, 427*(6976), 695. doi: 10.1038/427695a

Peer, D., Karp, J. M., Hong, S., Farokhzad, O. C., Margalit, R., & Langer, R. (2007). Nanocarriers as an emerging platform for cancer therapy. *Nat Nanotechnol, 2*(12), 751-760. doi: 10.1038/nnano.2007.387

Peppas, N. A. (2004). Intelligent therapeutics: biomimetic systems and nanotechnology in drug delivery. *Adv Drug Deliv Rev, 56*(11), 1529-1531. doi: 10.1016/j.addr.2004.07.001

Peracchia, M. T., Harnisch, S., Pinto-Alphandary, H., Gulik, A., Dedieu, J. C., Desmaele, D., et al. (1999). Visualization of in vitro protein-rejecting properties of PEGylated stealth polycyanoacrylate nanoparticles. *Biomaterials, 20*(14), 1269-1275.

Prabhakar, U., Maeda, H., Jain, R. K., Sevick-Muraca, E. M., Zamboni, W., Farokhzad, O. C., et al. (2013). Challenges and key considerations of the enhanced permeability and retention effect for nanomedicine drug delivery in oncology. *Cancer Res, 73*(8), 2412-2417. doi: 10.1158/0008-5472.CAN-12-4561

Pridgen, E. M., Langer, R., & Farokhzad, O. C. (2007). Biodegradable, polymeric nanoparticle delivery systems for cancer therapy. *Nanomedicine (Lond)*, *2*(5), 669–680. doi:10.2217/17435889.2.5.669 PMID:17976029

Quintana, A., Raczka, E., Piehler, L., Lee, I., Myc, A., & Majoros, I. et al. (2002). Design and function of a dendrimer-based therapeutic nanodevice targeted to tumor cells through the folate receptor. *Pharmaceutical Research*, *19*(9), 1310–1316. doi:10.1023/A:1020398624602 PMID:12403067

Rice, J. R., Gerberich, J. L., Nowotnik, D. P., & Howell, S. B. (2006). Preclinical efficacy and pharmacokinetics of AP5346, a novel diaminocyclohexane-platinum tumor-targeting drug delivery system. *Clin Cancer Res, 12*(7 Pt 1), 2248-2254. doi: 10.1158/1078-0432.CCR-05-2169

Riehemann, K., Schneider, S. W., Luger, T. A., Godin, B., Ferrari, M., & Fuchs, H. (2009). Nanomedicine--challenge and perspectives. *Angewandte Chemie (International ed. in English)*, *48*(5), 872–897. doi:10.1002/anie.200802585 PMID:19142939

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Rolland, J. P., Maynor, B. W., Euliss, L. E., Exner, A. E., Denison, G. M., & DeSimone, J. M. (2005). Direct fabrication and harvesting of monodisperse, shape-specific nanobiomaterials. *Journal of the American Chemical Society*, *127*(28), 10096–10100. doi:10.1021/ja051977c PMID:16011375

Rosen, H., & Abribat, T. (2005). The rise and rise of drug delivery. *Nat Rev Drug Discov, 4*(5), 381-385. doi: 10.1038/nrd1721

Ruggiero, A., Villa, C. H., Bander, E., Rey, D. A., Bergkvist, M., Batt, C. A., et al. (2010). Paradoxical glomerular filtration of carbon nanotubes. *Proc Natl Acad Sci U S A, 107*(27), 12369-12374. doi: 10.1073/pnas.0913667107

Saha, R. N., Vasanthakumar, S., Bende, G., & Snehalatha, M. (2010). Nanoparticulate drug delivery systems for cancer chemotherapy. *Molecular Membrane Biology*, *27*(7), 215–231. doi:10.3109/09687688.2010.510804 PMID:20939772

Schwartz, J. A., Shetty, A. M., Price, R. E., Stafford, R. J., Wang, J. C., Uthamanthil, R. K., et al. (2009). Feasibility study of particle-assisted laser ablation of brain tumors in orthotopic canine model. *Cancer Res, 69*(4), 1659-1667. doi: 10.1158/0008-5472.CAN-08-2535

Sengupta, S., Eavarone, D., Capila, I., Zhao, G., Watson, N., Kiziltepe, T., et al. (2005). Temporal targeting of tumour cells and neovasculature with a nanoscale delivery system. *Nature, 436*(7050), 568-572. doi: 10.1038/nature03794

Shi, J., Xiao, Z., Kamaly, N., & Farokhzad, O. C. (2011). Self-assembled targeted nanoparticles: Evolution of technologies and bench to bedside translation. *Accounts of Chemical Research*, *44*(10), 1123–1134. doi:10.1021/ar200054n PMID:21692448

Shukla, S., Ablack, A. L., Wen, A. M., Lee, K. L., Lewis, J. D., & Steinmetz, N. F. (2013). Increased tumor homing and tissue penetration of the filamentous plant viral nanoparticle Potato virus X. *Molecular Pharmaceutics*, *10*(1), 33–42. doi:10.1021/mp300240m PMID:22731633

Singer, J. W. (2005). Paclitaxel poliglumex (XYOTAX, CT-2103): a macromolecular taxane. *J Control Release, 109*(1-3), 120-126. doi: 10.1016/j.jconrel.2005.09.033

Singer, J. W., Baker, B., Vries, P., Kumar, A., Shaffer, S., & Vawter, E. et al. (2003). Poly-(L)-glutamic acid-paclitaxel (CT-2103) [XYOTAX], a biodegradable polymeric drug conjugate: characterization, preclinical pharmacology, and preliminary clinical data. *Advances in Experimental Medicine and Biology*, *519*, 81–99. doi:10.1007/0-306-47932-X_6 PMID:12675210

Sinha, R., Kim, G. J., Nie, S., & Shin, D. M. (2006). Nanotechnology in cancer therapeutics: bioconjugated nanoparticles for drug delivery. *Mol Cancer Ther, 5*(8), 1909-1917. doi: 10.1158/1535-7163.MCT-06-0141

Sood, P., Thurmond, K. B. II, Jacob, J. E., Waller, L. K., Silva, G. O., Stewart, D. R., & Nowotnik, D. P. (2006). Synthesis and characterization of AP5346, a novel polymer-linked diaminocyclohexyl platinum chemotherapeutic agent. *Bioconjugate Chemistry*, *17*(5), 1270–1279. doi:10.1021/bc0600517 PMID:16984138

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Sudimack, J. J., Guo, W., Tjarks, W., & Lee, R. J. (2002). A novel pH-sensitive liposome formulation containing oleyl alcohol. *Biochim Biophys Acta, 1564*(1), 31-37.

Swartz, M. A., & Fleury, M. E. (2007). Interstitial flow and its effects in soft tissues. *Annual Review of Biomedical Engineering*, *9*(1), 229–256. doi:10.1146/annurev.bioeng.9.060906.151850 PMID:17459001

Thevenot, J., Troutier, A. L., David, L., Delair, T., & Ladaviere, C. (2007). Steric stabilization of lipid/polymer particle assemblies by poly(ethylene glycol)-lipids. *Biomacromolecules*, *8*(11), 3651–3660. doi:10.1021/bm700753q PMID:17958441

Tribler, L., Jensen, L. T., Jorgensen, K., Brunner, N., Gelb, M. H., & Nielsen, H. J. et al. (2007). Increased expression and activity of group IIA and X secretory phospholipase A2 in peritumoral versus central colon carcinoma tissue. *Anticancer Research*, *27*(5A), 3179–3185. PMID:17970059

Tsukioka, Y., Matsumura, Y., Hamaguchi, T., Koike, H., Moriyasu, F., & Kakizoe, T. (2002). Pharmaceutical and biomedical differences between micellar doxorubicin (NK911) and liposomal doxorubicin (Doxil). *Japanese Journal of Cancer Research*, *93*(10), 1145–1153. PMID:12417045

Uchegbu, I. F. (2006). Pharmaceutical nanotechnology: Polymeric vesicles for drug and gene delivery. *Expert Opinion on Drug Delivery*, *3*(5), 629–640. doi:10.1517/17425247.3.5.629 PMID:16948558

Uchino, H., Matsumura, Y., Negishi, T., Koizumi, F., Hayashi, T., Honda, T., et al. (2005). Cisplatin-incorporating polymeric micelles (NC-6004) can reduce nephrotoxicity and neurotoxicity of cisplatin in rats. *Br J Cancer, 93*(6), 678-687. doi: 10.1038/sj.bjc.6602772

Ullal, A. V., Reiner, T., Yang, K. S., Gorbatov, R., Min, C., & Issadore, D. et al. (2011). Nanoparticle-mediated measurement of target-drug binding in cancer cells. *ACS Nano*, *5*(11), 9216–9224. doi:10.1021/nn203450p PMID:21962084

Valencia, P. M., Hanewich-Hollatz, M. H., Gao, W., Karim, F., Langer, R., Karnik, R., et al. (2011). Effects of ligands with different water solubilities on self-assembly and properties of targeted nanoparticles. *Biomaterials, 32*(26), 6226-6233. doi: 10.1016/j.biomaterials.2011.04.078

Vasey, P. A., Kaye, S. B., Morrison, R., Twelves, C., Wilson, P., & Duncan, R. et al. (1999). Phase I clinical and pharmacokinetic study of PK1 [N-(2-hydroxypropyl)methacrylamide copolymer doxorubicin]: First member of a new class of chemotherapeutic agents-drug-polymer conjugates. Cancer Research Campaign Phase I/II Committee. *Clinical Cancer Research*, *5*(1), 83–94. PMID:9918206

Venugopal, J., Prabhakaran, M. P., Low, S., Choon, A. T., Zhang, Y. Z., Deepika, G., & Ramakrishna, S. (2008). Nanotechnology for nanomedicine and delivery of drugs. *Current Pharmaceutical Design*, *14*(22), 2184–2200. doi:10.2174/138161208785740180 PMID:18781971

Vial, F., Rabhi, S., & Tribet, C. (2005). Association of octyl-modified poly(acrylic acid) onto unilamellar vesicles of lipids and kinetics of vesicle disruption. *Langmuir*, *21*(3), 853–862. doi:10.1021/la048039v PMID:15667160

Visaria, R., Bischof, J. C., Loren, M., Williams, B., Ebbini, E., Paciotti, G., et al. (2007). Nanotherapeutics for enhancing thermal therapy of cancer. *Int J Hyperthermia, 23*(6), 501-511. doi: 10.1080/02656730701611241

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Wang, J., Tian, S., Petros, R. A., Napier, M. E., & Desimone, J. M. (2010). The complex role of multivalency in nanoparticles targeting the transferrin receptor for cancer therapies. *Journal of the American Chemical Society*, *132*(32), 11306–11313. doi:10.1021/ja1043177 PMID:20698697

Yu, B., Tai, H. C., Xue, W., Lee, L. J., & Lee, R. J. (2010). Receptor-targeted nanocarriers for therapeutic delivery to cancer. *Molecular Membrane Biology*, *27*(7), 286–298. doi:10.3109/09687688.2010.521200 PMID:21028937

Zamboni, W. C. (2008). Concept and clinical evaluation of carrier-mediated anticancer agents. *Oncologist, 13*(3), 248-260. doi: 10.1634/theoncologist.2007-0180

Zetter, B. R. (2008). The scientific contributions of M. Judah Folkman to cancer research. *Nat Rev Cancer, 8*(8), 647-654. doi: 10.1038/nrc2458

Zhang, L., Chan, J. M., Gu, F. X., Rhee, J. W., Wang, A. Z., Radovic-Moreno, A. F., et al. (2008). Self-assembled lipid--polymer hybrid nanoparticles: a robust drug delivery platform. *ACS Nano, 2*(8), 1696-1702. doi: [pii]10.1021/nn800275r 10.1021/nn800275r

Zhang, X. Q., Xu, X., Bertrand, N., Pridgen, E., Swami, A., & Farokhzad, O. C. (2012). Interactions of nanomaterials and biological systems: Implications to personalized nanomedicine. *Adv Drug Deliv Rev, 64*(13), 1363-1384. doi: 10.1016/j.addr.2012.08.005

KEY TERMS AND DEFINITIONS

ATI-1123: A liposomal formulation of docetaxel.
CDP: Cyclodextrin-based polymer.
ECM: Extracellular matrix.
EPR: Enhanced permeability and retention.
GAGs: Glycosaminoglycans.
NP: Nanoparticles.
PC-3: Human prostate carcinoma cells.
PEG: Polyethylene glycol.
PLA: Poly (lactic acid).
PTX: Paclitaxel.
SEM: Scanning electron microscope.
Tf: Transferrin.

This work was previously published in the Handbook of Research on Diverse Applications of Nanotechnology in Biomedicine, Chemistry, and Engineering edited by Shivani Soni, Amandeep Salhotra, and Mrutyunjay Suar, pages 167-191, copyright year 2015 by Engineering Science Reference (an imprint of IGI Global).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 15
QSAR of Antioxidants

Omar Deeb
Al-Quds University, Palestine

Mohammad Goodarzi
Katholieke Universiteit Leuven, Belgium

ABSTRACT

Antioxidants are substances that protect cells from the damaging effects of oxygen radicals, which are chemicals that play a part in some diseases such as cancer and others. Antioxidants are expected to be promising drugs in the management of these diseases by removing oxidative stress. Most of the modeling approaches involved in designing new antioxidants is based on Quantitative Structure-Activity Relationship (QSAR). A number of QSAR studies have been conducted to elucidate the structural requirements of antioxidants for their activities in order to predict the potency of these compounds with regard to the targeted activity and to direct the synthesis of more potent analogues. The main focus of this chapter is on the QSAR modeling of antioxidant compounds. The authors provide different QSAR studies of antioxidant compounds and try to compare between them in terms of the best models obtained and their use in designing potential new drugs.

INTRODUCTION

Generally, antioxidants are defined as molecules, which are capable of slowing down the oxidation mechanism of an oxidizable compound. At the same time, oxidation represents an essential part of aerobic life and our metabolism, since oxygen is the ultimate electron acceptor. Antioxidants play a significant role in cell protection against oxidative stress. Because of the ability of antioxidants, scientific evidence suggests that these molecules reduce the risk of chronic diseases including cancer and heart disease. Primary sources of naturally occurring antioxidants are whole grains, fruits, vegetables etc.

Therefore, much attention has been devoted recently to research into the role of plant-derived antioxidants in food and human health. The beneficial influence of many foodstuffs and beverages including fruits, vegetables, tea, coffee, and cacao on human health has been recently recognized to originate from their antioxidant activity.

DOI: 10.4018/978-1-5225-0549-5.ch015

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

In spite of the fact that there are many experimental techniques used to determine antioxidant activity of a given molecule, such techniques are not only time consuming but also costly. Alternatively, using Quantitative Structure-Activity Relationship method, one can screen a large number of compounds and check for their potency as an antioxidant. For this purpose, the most common methods used *in vitro* determination of antioxidant activity are reviewed and presented.

This chapter deals with Quantitative Structure-Activity Relationship (QSAR) of antioxidants. At first, the definition of an antioxidant is given including it's mechanism of action and methods of antioxidant determination. The issue of computer aided drug design is then discussed, focusing on the quantitative structure activity relationship (QSAR). The QSAR section starts with history and the earliest efforts made in this field, types of descriptors, statistical analysis methods from a simple linear regression such as Multiple Linear Regression (MLR) to biased regression based on Principal Component Analysis (PCA), such as Partial Least Squares (PLS) to nonlinear techniques such as Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs). Moreover, a part of this chapter is devoted to the validation of QSAR models, explaining the internal and external validations. At the end, different QSAR studies of the antioxidant compounds are discussed. These QSAR models were built on different types of compounds such as Flavonoids, Hydroxyphenylureas, curcumin analogues, Hydroxybenzalacetones, Phenolic derivatives bearing NO donor groups and Coumarin derivatives. Finally, it was concluded that more attention should go to the building of QSAR models for antioxidant activities in order to design more efficient drugs with potential antioxidant activity.

BACKGROUND

Antioxidants

Prior to defining what an antioxidant is, we would like to use the term oxidation as a chemical reaction which transfers electrons or hydrogen from one substance to another. These oxidation reactions can produce free radicals, which are unstable molecules that lose one of their electrons and therefore become unbalanced and highly reactive. These unstable molecules can then start chain reactions in a cell and therefore, may cause damage or result in death of these cells. These free radicals are free moving compounds that travel around the body trying to become stable by taking electrons from healthy compounds. Further free radicals are formed after a successful steal, damaging healthy cells in the progression. These free radicals are the main reason behind heart disease, cancer, aging and other diseases. Therefore, an antioxidant is a substance that inhibits the oxidation of other molecules and terminates the chain reactions by removing the free radicals and inhibiting other oxidation reactions. In other words, an antioxidant is a stable molecule that donates an electron to a free radical neutralizing it and therefore, reducing its capability to damage other cells.

Antioxidants are extensively used as ingredients in dietary supplements with the hope of maintaining health and preventing diseases. At the same time, these antioxidants have many industrial uses such as preservatives in food and cosmetics, and preventing the degradation of rubber. As is known, antioxidants are abundant in fruits and vegetables. Examples of antioxidants that are formed during metabolism in the body, or endogenous antioxidants (Enzymes), are glutathione, ubiquinol, uric acid etc. Some antioxidants are present in the diet while others must be supplied in the diet such as vitamin C, vitamin E and others. In this case, they are called exogenous antioxidants and can be derived from natural sources

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

but also can be synthesized compounds (Benzie, 2003; Knight, 1998; Matill, 1947; Sies, 1997; Valko, Leibfritz, Moncol, Cronin, Mazur & Telser, 2007; Vertuani, Angusti & Manfredini, 2004; Wolf, 2005)

Recently, antioxidants have attracted extensive attention in relation to radicals and oxidative stress, cancer prophylaxis and therapy. Phenols and polyphenols are the target analysis in numerous such cases; they may be detected by enzymes like tyrosinase or other phenol oxidases, or even by plant tissues containing these enzymes. The recommendations based on epidemiological studies are that fruits, vegetables and less processed fast foods guarantee the best defense against the increase of diseases caused by oxidative stress, such as cancer, hypertension and others. The elucidation consists in the valuable health effect, owing to antioxidants present in fruit and vegetables. There are numerous antioxidants in dietary plants such as: phenolic compounds, benzoic acid derivatives and flavonoids. (Halvorsen, Carlsen, Phillips, Bohn & Holte, 2006; Halvorsen, Holte, Myhrstad, Barikmo & Hvattum, 2002)

MECHANISM OF ACTION OF ANTIOXIDANTS

Two main mechanisms have been proposed regarding the mechanism of action of antioxidants. The first one is related to chain breaking in which the main antioxidant donates an electron to the free radical, which exists in the system, while the second mechanism involves the removal of reactive oxygen/nitrogen species initiators (ROS/RNS). The mechanism of action can be summarized as: Initiation, Propagation, Branching and Termination (Antolovic, Prenzler, Patsalides, McDonald & Robards, 2002)

The antioxidants in the defense systems act at different levels: preventive antioxidants which suppress the formation of free radicals, antioxidants that scavenge the active radicals to suppress chain initiation or breaking, the repair and de novo antioxidants.

In the initiation step, a substrate molecule (SH) such as lipid reacts with the initiating oxidizing radical (R.) and a highly reactive allyl radical (S.) is produced as in the following equation:

$$SH + R. \rightarrow S. + RH \tag{1}$$

In the propagation step, the allyl radical (S.) reacts quickly with oxygen to produce a lipid peroxyl radical (SOO.). This allyl peroxyl radical (SOO.) can further oxidize the substrate molecule producing lipid hydroperoxides (SOOH) as shown in the following Equations (2 and 3:

$$S. + O_2 \rightarrow SOO. \tag{2}$$

$$SOO. + SH \rightarrow S. + SOOH \tag{3}$$

In the branching step, the lipid hydroperoxides (SOOH) can break down to alcohols, aldehydes, ketones and others including alkoxyl radical (SO.), which can be seen in the following Equations (4 and 5:

$$SOOH \rightarrow SO. + HO. \tag{4}$$

$$2SOOH \rightarrow SOO. + SO. + H_2O \tag{5}$$

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

In the termination step, combination of radicals will produce non-radical products as is shown below:

$$SO.+SO.$$

$$SOO.+SOO.$$

$$SO.+SOO.$$

The primary antioxidant (AntiH) delay or inhibit the initiation step by reacting with a substrate radical or inhibit the propagation step by reacting with peroxyl or alkoxyl radicals as shown in the following Equations (6, 7 and 8:

$$S.+AntiH \rightarrow SH + Anti. \tag{6}$$

$$SOO.+AntiH \rightarrow SOOH + Anti. \tag{7}$$

$$SO.+AntiH \rightarrow SOH + Anti. \tag{8}$$

While the secondary or preventive antioxidant slow down, the speed of oxidation by taking away from the substrate or singlet oxygen increases.

In order to illustrate the mechanism of action of antioxidants, in the initiation step (Equation 1), the flavonoids (ROH) is as a substrate that can react with initiating oxidizing free radical (DPPH.) to produce flavonoid phenoxyl radical (RO.) and DPPHH. Another example is the reaction of a reduced glutathione molecule (GSH) with hydroxyl radical (OH.), which produces glutathione radical (GS.) and a water molecule (H_2O). Vitamine C is also another example, Ascorbate $AscH^-$ for instance can donate a hydrogen atom to a free radical molecule (R.) to neutralize it and produce an ascorbate radical $Asc.^-$

METHODS OF ANTIOXIDANTS DETERMINATION

There are various analytical techniques used to determine total antioxidant activity. These techniques are mainly based on spectrometric, electrochemical and Chromatographic techniques. The spectrometric techniques depend on the reaction of a radical, radical cation or complex with an antioxidant molecule capable of donating a hydrogen atom. An example of the spectrometric methods is the DPPH method. The DPPH (2,2-diphenyl-1-picrylhydrazyl) is a stable free radical because of the delocalization of the spare electron on the whole molecule. This delocalization determines the occurrence of purple color at an absorption band of 520nm. When DPPH reacts with a hydrogen donor, a molecular form of DPPH is formed accompanied with the vanishing of the color. So the absorbance decrease depends linearly on the antioxidant concentration (Pisoschi, Cheregi & Danet, 2009; Thaipong, Boonprakob, Crosby, Cisneros-Zevallos & Byrne, 2006)

On the other hand, electrochemical techniques such as cyclic voltammetry, amperometric and biamperometric methods are also used for antioxidant activity determination. As an example, in the Amperometric method, it involves measurement of the intensity of the current that flows between a working electrode and a reference electrode at a fixed applied potential. The amperometric determina-

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

tion of the antioxidant activity based on the reduction of DPPH (2,2-diphenyl-1-picrylhydrazyl) at the carbon electrode (Milardovic, Ivekovic & Grabaric, 2006)

And finally the other popular analytical technique, the Chromatographic technique that is applied to antioxidant separation and detection. The chromatographic techniques are used before spectrometric or electrochemical assessment of the total antioxidant capacity. Examples of the chromatographic methods used are Gas Chromatography (GC) and High Performance Liquid Chromatography (HPLC). For example, the gas chromatographic method is used for separating and analyzing compounds that can be vaporized without decomposition. The separation process is carried out between two phases, a liquid stationary phase, which is a liquid on an inert solid support, and a gas mobile phase such as helium. In GC, a comparison of retention times took place. An example of the detectors used in GC is the flame ionization detector. For an antioxidant capacity determination using GC, the turmeric oil was determined by gas chromatography with a flame ionization detector and it was coupled with mass spectrometry. The testing of turmeric oil and its fractions for antioxidant activity was done using a carotene-linoleate model and the phosphomolybdenum method (Jayaprakasha, Jena, Negi & Sakariah, 2002).

COMPUTER AIDED DRUG DESIGN (CADD)

The most important challenges that medicinal chemists face today is the design of new drugs with improved potency and less side-effects for treating human diseases such as AIDS, malaria and others. Medicinal chemists begin the process by taking a lead structure and then finding analogs exhibiting the preferred biological activities. Next, they use their experience and chemical insight to eventually choose a candidate analog for further development. This process was not only difficult but also expensive and time consuming. The conventional methods of drug discovery are now being supplemented by shorter approaches made possible by the accepting of the molecular processes involved in the original disease. In this view, the preliminary point in drug design is the molecular target, which is a receptor or enzyme in the body as an option of the existence of the known lead structure.

Samples of protein targets were isolated and X-ray crystallography discovered their molecular structural design. Molecules are conceived either on the basis of similarities with recognized reference structures or on the basis of their complementarily with the three dimensional (3D) structure of well-known active sites.

The techniques currently on hand provide widespread insight into exact molecular features that are in charge for the regulation of biological processes: molecular geometries, electronic features and others. All these structural characteristics are of crucial importance in the understanding of Structure-Activity Relationships and in rational drug design.

Rational drug design is based on the belief that the biological properties of drugs are related to their actual structural features. What has changed along the years is the way molecules are perceived and defined. In the past, medicinal chemists considered molecules as simple two-dimensional (2D) entities with related chemical and physicochemical properties. Quantitative structure-activity relationships (QSAR) concepts began to be considered and became well accepted.

However, most of these properties have not been well represented by the basic numerical parameters considered to characterize these features: the interactions between a ligand and a protein require much more information than that included in substituent indices characterizing the molecular properties. Now, it has shown that consideration of the full detailed properties of 3D structures is necessary which allow us to understate stereochemical features.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

The effective design of chemical structures with the desirable therapeutic properties is directed towards Computer Aided-Drug Design (CADD) a well-established area of Computer Aided-Molecular Design (CAMD). These techniques cover new methodologies, such as molecular modeling and QSAR. Molecular modeling can be simply considered as a range of automated techniques based on theoretical chemistry methods and experimental data that can be used to predict molecular and biological properties. QSAR will be discussed in more details in the next section.

The main applications of CAMD are the clarification of the basic requirements for a compound to obtain a determined activity, the simulation of the binding between a ligand and the receptor, the discovery of new active compounds and the prediction of activities for non-synthesized analogues. These applications convert CAMD to be used in drug design.

Computer Aided-Molecular Design (CAMD) is predicted to contribute to the discovery of "brilliant" molecules conceived on the basis of exact three-dimensional details. Two major modeling strategies right now are used in the designing of new drugs. In the first strategy, the three-dimensional features of a known receptor site are directly considered whereas in the second strategy, the design is based on the comparative analysis of the structural features of known active and inactive molecules that are interpreted in terms of their complementarily with a supposed receptor site model.

The improvements in computer speed and capacity increased the number of lead compounds available for further research. However, not only the number of feasible drug candidates increased, but also the costs and time devoted in various drug discovery processes was reduced, improving the effectiveness of the drug development.

One of the initial approaches to decrease these costs were attempted by correlating the biological function of a compound with its chemical structure, expressed in terms of molecular structural descriptors, by means of QSAR techniques. QSAR believe that the biological activity of a compound is a result of its chemical structure. Within the QSAR approach, the descriptor variables are not physically measured but computed, therefore, they are easy and cheap to generate even for large molecular sets. This discipline was promoted by Hansch and his group (Fujita, 1990). It was based on the determination of mathematical equations expressing the biological activities as a function of molecular parameters.

QUANTITATIVE STRUCTURE ACTIVITY RELATIONSHIPS (QSAR)

QSAR is a way of finding and modeling a simple mathematical equation between molecular descriptors calculated from structures of a set of molecules with their reactivity in order to be able to predict the reactivity of unknown compounds.

Applications of QSAR can be extended to any molecular design purpose, including prediction of different kinds of biological activities, lead compound optimization and prediction of novel structural leads in drug discovery. The process of building a QSAR model is similar, apart from what type of property is being predicted. It consists of several steps which hopefully lead to the design of new compounds with the desired activity profile.

The first step of building a QSAR model is to select a training set consisting of compounds with their experimental biological activities. Ideally, each of these biological activities should cover the range of possible values for that activity. The next step is to optimize the molecular structure by a proper technique. Afterwards, one should compute descriptors that contain sufficient relevant information about the biological phenomenon. However, it is difficult to predict in advance which descriptor is informative

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

and which one is not. Once descriptors have been calculated, it is necessary to pick which should be included in the QSAR model. Therefore, feature selection techniques are used to select the most relevant descriptors from a pool of descriptors. One way is simply to use the correlation coefficient, which gives a quantitative measure of how well each descriptor describes the activity. Thus, the descriptor with the highest correlation coefficient can be selected. Next step, a data analysis is needed to calculate the best mathematical expression linking together the descriptors and biological activities, in which information relating the essential features of the chemical and biological data structure is obtained. In the final step, validation and predictions for non-tested compounds will take place. However, the predictive capability of the model first is verified experimentally (using an external set of compounds which are not used in calibration). This is talented by biological testing of some additional compounds (test set) in the same way as the training set and then comparing the experimental finding with the values predicted by the QSAR model. If the QSAR predicts within acceptable restrictions, it may be used for a more extensive prediction of more compounds.

History of QSAR

Crum-Brown and Fraser expressed the suggestion that the physiological action of a substance was a function of its chemical composition. Later, in 1893, Richet showed that the cytotoxicities of a dissimilar set of uncomplicated organic compounds were inversely related to their corresponding water solubility. After that, Meyer and Overton independently recommended that the narcotic action of a group of organic molecules correlated with their olive oil/water partition coefficients. The extensive work of Albert, and Bell and Roblin established the importance of ionization of bases and weak acids in bacteriostatic activity. In the physical organic border, great progress was made in the clarification of substituent effects on organic reactions, led by the influential job of Hammett. Taft invented a way for separating polar, steric, and resonance effects and introducing the first steric parameter, *ES.*

The contributions of Hammett and Taft together laid the mechanistic source for the progress of the QSAR model by Hansch and Fujita. The researchers (Hansch, Maloney, Fujita & Muir, 1962) published their bright study on the Structure-Activity Relationships of plant growth regulators and their dependency on Hammett constants and hydrophobicity. A Linear Free Energy Relationships (LFER) related model published by the researchers (Fujita, Iwasa & Hansch, 1964a; Hansch & Fujita, 1964b) considered to be the official beginning for QSAR. Their fragment and additive group contribution idea added two things: the use of calculated properties to correlate with biological activities and the detection that multiple properties may influence the biological activity. For this purpose, they implemented the use of the computer to fit QSAR equations.

The so-called Hansch equation (Hansch, 1969) was developed to correlate physicochemical properties (descriptors) with biological activities and is given in a general form by the following equation (9):

$$\log \frac{1}{C} = a(\log P)^2 + b\log P + cs + \cdots k \tag{9}$$

where *C* is the molar concentration that produces the biological effect; *P* is the octanol/water partition coefficient and *s* is the electronic Hammett constant.

Besides the Hansch approach, other methodologies were also developed to deal with Structure-Activity questions. The Free-Wilson approach (Free & Wilson, 1964) addresses Structure-Activity studies in a

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

congeneric series in which the contribution of each structural feature was a parameter of interest. These parameters, also called indicator variables, codify the presence or absence of a particular structural feature.

Descriptors

A common question in QSAR is how to describe molecules and their physicochemical properties (descriptors). The nature of the descriptors used and the extent to which they instruct the structural properties related to the biological activity is a critical step in a QSAR study (Downs, 2004). It has been estimated that over a thousand of molecular descriptors can now be calculated from a given molecular structure (Devillers & Balaban, 1999; Karelson, 2000; Todeschini, Consonni & Pavan, 2001). Most of those descriptors can be calculated by using commercial software packages such as CODESSA (Katritzky, Lobanov & Karelson, 1994), DRAGON (Todeschini, Consonni & Pavan, 2001) and others. The various descriptors in use can be largely categorized as being constitutional, topological, electrostatic, geometrical, or quantum chemical.

Constitutional descriptors give a simple description of what is in the molecule. For example, the number of heteroatoms, the number of rings, the number of double bonds, etc. can be categorized as constitutional descriptors type. Constitutional descriptors often appear in a QSAR equation when the property being predicted varies with the size of the molecule.

Topological descriptors are numbers that give information about the bonding collection in a molecule. They are derived from graph representation of chemical structures; they attempt to encode the size, shape, or branching in the compound by the handling of graph-theoretical aspects of the structures (Silipo & Vittoria, 1990). Some examples are Randic indices, Kier and Hall indices, Weiner index (sum of the chemical bonds existing between all pairs of heavy atoms in the molecule), the connectivity index and others.

Electrostatic descriptors are single values that give information about the molecular charge division. Some examples are polarity indices and polarizability. One of the most commonly used electrostatic descriptors is the topological polar surface area (TPSA), which gives an indication of the portion of the molecular surface composed of polar groups against nonpolar groups. Another deeply used descriptor is the octanol–water partition coefficient, which is designated by a specific prediction scheme such as ClogP or MlogP.

Geometrical descriptors are single values that describe the molecule's size and shape as well as the degree of complementarity of a ligand and the receptor. They are developed from three-dimensional models of molecules, and derived from molecular surface area calculations. Some examples are moments of inertia, molecular volume molecular surface area, and other parameters that describe length, height, and width.

Quantum chemical descriptors give information about the electronic structure of the molecule. These types of descriptors are calculated based on molecular orbital calculations and they mainly describe electronic interaction. This includes for example, the energy of the highest occupied molecular orbital, E_{HOMO}, which is a quantitative measure for the chemical reactivity of the compound-ionization potential of a molecule, the energy of the lowest unoccupied molecular orbital, E_{LUMO}, which accounts for the electron affinity, refractivity and total energy. The E_{HOMO}–E_{LUMO} gap or ionization potential can be important descriptors for predicting how molecules will react.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

New nodal angle quantum descriptors - the Frontier Orbital Phase Angles - suggested by Clare (1998) which are considered as novel QSAR descriptors for benzene derivatives will be discussed in the application part, the QSAR of antioxidants and especially flavonoids.

STATISTICAL METHODS

Chemometrics methods are the mathematical basis for the development of QSAR models. Chemometric methods (Eriksso, Johansson, Kettaneh-Wold & Wold, 2001) are used to extract information from QSAR data using tools of statistics, mathematics and informatics. The applications of these methods are combined with the important goal of explanation and prediction of non-synthesized test compounds. Many different statistical methods are available in the literature and the selection of the appropriate method is critical (Xu & Zhang, 2001). However, we will try to explain some basic techniques, which were extensively used in QSAR/QSPR studies.

- **Multiple Linear Regression (MLR):** (Montgomery & Peck, 1992) can be considered as an easy interpretable linear regression method. Regression analysis correlates independent X variables or descriptors (physicochemical parameters) with dependent Y variables (biological data). The regression model assumes a linear relationship between *m* molecular descriptors and the response (biological activity) variable. This relationship can be expressed with the single multiple-term linear equation as follows:

$$y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_m x_m + e\,. \tag{10}$$

  The MLR analysis calculates the regression coefficients, b_i, by minimizing the residuals, **e**, which quantify the deviations between the data (Y) and the model (Y'), as in the case of simple linear regression. However, one cannot build a model when the number of variables is higher or equal to the number of compounds. On the other hand, multicollinearity is another problem of MLR, which encourage researcher to use biased regression such as Principal Component Regression or Partial Least Squares techniques. These techniques reduce the dimension of the data set by applying a decomposition strategy (e.g. Principal Component Analysis).
- **Principal Component Analysis (PCA):** As we mentioned above, one way to reduce the dimensionality of the data set descriptors (X matrix) is the so-called Principal Component Analysis (PCA) technique (Jolliffe, 1986). It seeks to find a new set of variables named Principal Components (PC) showing the data in order of decreasing variance with the aim to state the main information in the variables by the principal components of X. The primary Principal Component (PC1) describes the maximum deviation in the whole data set. The subsequent principal component (PC2) describes the maximum remaining variance, and so forth, with each axis linearly independent, to the preceding axis. Some of the last components may be discarded to decrease the size of the model and stay away from over-fitting.
- **Partial Least Squares (PLS)** (Wold, Johansson & Cocchi, 1993) A biased regression technique used for describing a given response (dependent variable) as a function of a few latent variables (LVs), which are called PLS factors. It projects multivariate data into a space of lower size, and

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

certainly providing insight to see and model huge sets of data. The Partial Least Squares (PLS) regression method carries out regression using Latent Variables (LVs) from the independent and dependent data that are along their axes of most variation and are highly correlated. It is applied when the numbers of independent variables are more than the number of observations. Under these circumstances, it gives a stronger QSAR equation than Multiple Linear Regressions. Thus, PLS is able to examine complex structure-activity problems and to examine data in a more realistic way. The theory of PLS is extensively described in many research papers and therefore, we do not provide it here.

- **Principal Components Regression (PCR):** The Principal Components Regression (PCR) method uses linear regression to generate a model by means of the principal components as independent descriptors. PCR applies the scores from PCA as regressors in the QSAR model. Therefore, a multiple-term linear equation is generated and derived from a principal components analysis transformation of the independent variables.
- **Artificial Neural Networks (ANN) Method:** (Duprat, Huynh & Dreyfus, 1998; Novi, Nikolovska-Coleska & Solmajer, 1997; Tetko, Alessandro, Villa & Livingstone, 1996;) A non-linear technique inspired in the human brain, composed of many simple processing units called neurons. This method is also recognized as learning algorithms. The aim is to simulate the various shells of the neurons, where each neuron is connected to a number of neighboring neurons with variable coefficients of connectivity that signify the strength of these associations. The learning process consists of adjusting the coefficient so that the network provided as an output of the suitable results. In neural networks, a training set is used to train the network, and then the network is used to predict the property (biological activity) that it was trained to predict. This technique can be associated with principal components analysis in which it is referred as 'PC-ANN.'
- **Support Vector Machine (SVM):** Can be applied to regression by the introduction of an alternative loss function. In support vector regression (SVR) (Gunn, 1997) the basic idea is to map the data X into a higher-dimensional feature space via a nonlinear mapping and then perform a linear or nonlinear regression in that space. Therefore, regression approximation addresses the problem of estimating a function based on a given data set.

VALIDATION OF QSAR MODELS

After the model equation is obtained, or the stability and the accuracy of the fit of the model, it is also significant to estimate the power and the validity of the model before using it to predict the biological activity. Validity is to establish the reliability and significance of the method for a particular use. For that reason, validation of a QSAR model must be done. There are two validation methods used for a QSAR model: internal and external validation techniques to establish the confidence and strength of the model.

Internal Validation

Internal validation uses the data set from which the model is built and checks for internal stability. Cross-Validation (CV) technique is widely employed as an internal validation method of statistical models (Allen, 1974; Wold, 1978a; 1991b). Usually, one compound of the calibration set is eliminated each time, then the model is recalculated using a training set the *n-1* (where n is number of compounds) remaining

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

compounds, so that the biological activity value for the extracted compound is predicted once for all compounds. This process is repeated *n* times for all the compounds of the calibration set, thus obtaining a prediction for each object. This process is referred to as leave-one-out (LOO) method. Also an alternative method can be defined when leaving out more than a compound of the data set at each time. This method is called leave n-out or Leave-many-Out (LMO) CV method or sometimes it is referred as leave-group-out (LGO). Calculation of the Squared Correlation Coefficient of the Cross-Validation procedure, that is, the coefficient of prediction q^2 must be done and it is by definition smaller than or equal to the overall r^2 (squared correlation coefficient) for a QSAR model. It is used as an investigative tool to estimate the predictive power of an equation obtained by using a regression method. Another procedure to test the validity of the model is the randomization test. Even with a huge number of compounds and a small number of descriptors, an equation can still have very poor predictive power. One way to test for this is by randomization of the compounds. The set of biological activity values is re-assigned arbitrarily to different compounds, and a new regression is done. This process is repeated many times. If the random models' biological activity prediction is analogous to the original equation within a given estimated self-confidence level, this means that the original model was obtained by chance. The random test analyzes the ability of the model to derive actual Structure-Activity relationships.

External Validation

A QSAR model with excellent goodness of fit and acceptable predictions may be deficient in real relationship between structural descriptors and biological activity. The perfect validity of the model is examined by external validation, which evaluates how well a built model is generalized. If a sufficiently large series of compounds with known activity are obtainable, the original data set can be divided into two subgroups, the training set and the test set. The training or calibration set is used to derive a calibration model that will be used later to predict the activities of the test or validation set compounds. In simple words, the external test set must not have been included in any stage of the building of the model to be used as test set.

The obtained predictions of the new generated model for the test set determine the validity of the model. The parameters quantifying the superiority of prediction of the external test set may be the same used as for the internal validation. The Sum of Squares Prediction Errors (SSPE) is widely used to account for the inconsistency.

QSAR of Antioxidants

Many QSAR papers have been published in the past two decades related to antioxidant activity. In this section we will report some QSAR case studies related to flavonoids, hydroxyphenylureas, curcumin analogues, hydroxybenzalacetones, phenolic derivatives bearing NO donor groups and coumarin derivatives.

Falvonoids

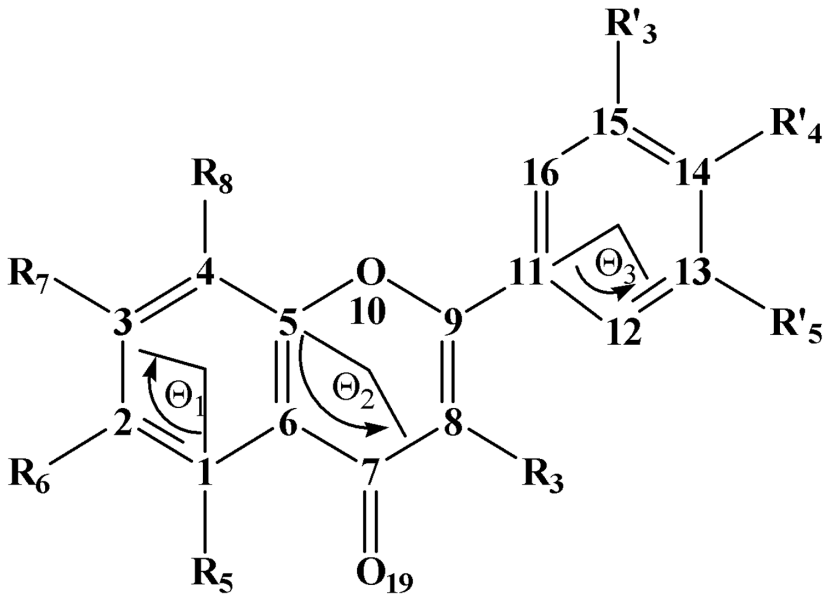
Flavonoids are a group of naturally occurring compounds present in fruits, vegetables, and beverages such as wine, tea and others. These flavonoids are a group of low molecular weight plant products, based on the parent compound, flavone (2-phenylchromone) and have shown potential for application in a variety of pharmacological targets. A large number of natural and synthetic flavonoids are being tested

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

as specific inhibitors of protein tyrosine kinase (PTK). The flavonoid-inhibitory activity is expressed as log $1/IC_{50}$, which is the molar concentration of the flavonoid necessary to give half-maximal inhibition compared to the control assay carried out in the absence of inhibitor, but in the presence of dimethyl sulphoxide carrier. Flavonoids have gained enormous interest because of their beneficial health effects such as anti-inflammatory, cardio-protective and anticancer activities. These findings have contributed to the remarkable increase in the expenditure and use of dietary supplements containing high concentrations of plant flavonoids. The pharmacological outcome of flavonoids is mainly because of their antioxidant activity and their inhibition of certain enzymes. A review paper has been published related to SAR and QSAR of the antioxidant activity of flavoniods (Dragan, Dusanka, Drago, Vesna, Bono & Nenad, 2007).

In the QSAR analysis performed by Deeb and Clare (Deeb & Clare, 2007). they have investigated the flavonoid-inhibitory activity of 54 analogues using the nodal angle descriptors (Clare, 2000; Clare & Supuran, 2005) and flipstep regression analysis described elsewhere (Clare, 2000; 2001a). For the flavonoid, calculating the angles in the three rings can be accomplished by entering the atom as numbered in Figure 1. The three rings are 6-membered rings numbered 1–6, 5–10 and 11-16 for rings 1, 2 and 3 respectively. The angles calculated by NODANGLE (Clare & Supuran, 2005) are then Θ_1, Θ_2 and Θ_3 as shown in Figure 1, measured at atoms 1, 5 and 11 respectively. A problem arises from the symmetry of the parent molecule; therefore, the program FLIPSTEP, a component of the MARTHA statistical package (Clare & Supuran, 2005a) was used. This set of flavonoids separates into two parts (symmetry wise): the chromone moiety and the phenyl ring. The chromone ring system has no vertical mirror planes or axes. Hence, ring 1 cannot be flipped into ring 2. Thus for this part of the molecule flip regression is not applied. The phenyl ring has C_{2v} symmetry, so flip regression is applicable to this. Thus, only ring 3 should be flipped. Figure 1 shows the numbering system of flavonoids.

Figure 1. Numbering of flavonoids skeleton used in MOPAC and NODANGLE calculations and angles used in the interpretations; the angle subscript indicates the ring number.



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

In the study carried out by Deeb and Clare (Deeb & Clare, 2007), it was demonstrated that the charge on O_{10} proved to be the most important factor. Low charge on O_{10} was found to be favorable to high activity. Furthermore, it was found that the charge on C_7 and the mean of absolute charge are significant variables. Moreover, it was shown that the orientation of the nodes on ring 3 are significant factors which indicate the importance of the electrostatic and quantum chemical descriptors for the interaction of flavonoids with the specific enzymatic active site playing an important role. Exactly which rings are involved becomes clear from the identity of the descriptors included in the regression model equation. They found that a model could be built using several molecular descriptors such as HOP1, SHOP1, HOP3, S2Θ1H, C2Θ1H, Vol, $QC_{7,}$ $QO_{10,}$ $Q_{mean,}$ C2Θ3H, S2Θ3H, C4Θ3L and S4Θ3L.

Where HOP1 is the highest occupied π orbital on ring 1, SHOP1 is the second highest occupied π orbital on ring 1, HOP3 is the highest occupied π orbital on ring 3, S2Θ1H is sin(2× the nodal angle in the highest occupied π orbital in ring 1), C2Θ1H is cos(2× the nodal angle in the highest occupied π orbital in ring 1), Vol is molecule volume, QC_7 is charge on $C_{7,}$ QO_{10} is charge on O_{10}, Q_{mean} is the mean absolute Mulliken charge, C2Θ3H is cos(2× the nodal angle in the highest occupied π orbital in ring 3), S2Θ3H is sin(2× the nodal angle in the highest occupied π orbital in ring 3), C4Θ3L is cos(4× the nodal angle in the lowest unoccupied π orbital in ring 3), S4Θ3L is sin (4× the nodal angle in the lowest unoccupied π orbital in ring 3).

The work of Deeb and Clare (Deeb & Clare, 2007) demonstrated that the nodal orientation terms have a powerful explanatory importance. They account for more of the variance in activity than is possible using the classical descriptors alone. However, a combination of the classical descriptors and the nodal orientation term gives even better explanatory of activity of the flavone analogues. The chromone moiety of the flavonoid structure is envisaged to be a mixed region for hydrophobic and electronic interactions, while the phenyl ring moiety, especially the substituents at the 3' and 4' position, are involved in electronic interactions with the enzyme. S4Θ3L, that is cos (4 × the nodal angle in the lowest unoccupied π orbital in ring 3), was identified to be an important descriptor. They used two relatively new techniques in this study. The first is flip regression, for handling the symmetry of the phenyl ring system. The second is the use of the orbital nodal angle descriptors. This is based on the concept that the stability of stacked aromatic systems is highly orientation-dependent. It is also dependent on the energies of those orbitals in the two aromatic systems that resemble the degenerate HOMO and LUMO of benzene. Adding classical variables improves the correlation. The classical variables found to be significant e.g. the charges on O10, C7, and charge polarization. Low charge on O10 was found to be favorable to high activity.

In a recent QSAR analysis, Deeb et al. (Deeb, Shaik & Agrawal, 2014) studied the interaction of flavonoids with GABA(A) receptor using different statistical methods. The QSAR models for binding affinity constants (log Ki) of 78 flavonoid ligands towards the benzodiazepine site of GABA (A) receptor complex were calculated using the machine learning methods: Artificial Neural Network (ANN) and Support Vector Machine (SVM) techniques. The models obtained were compared with those obtained using Multiple Linear Regression (MLR) analysis. The descriptor selection and model building were performed with 10 fold cross validation using the training data set. The SVM and MLR coefficient of determination values are 0.944 and 0.879, respectively for the training set, and are higher than those of ANN models. Though the SVM model shows improvement of training set fitting, the ANN model was superior to SVM and MLR in predicting the test set. The proposed models could identify and provide some important information, which is responsible for drug binding affinity. Employed randomization test indicates that the models obtained in this study are better than those obtained by chance. These models could be used for designing new drugs.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

The model equation represents the best MLR model obtained in this study (Deeb, Shaik & Agrawal, 2014). The selected descriptors for this study were MSD, PJI2, IVDE, MATS8v, MATS2e, GATS8e, JGI9, L/Bw, R4e+ and nRORPh. A brief description of these descriptors are as follows:

- **MSD (Topological Descriptor):** Mean Square Index (Balaban).
- **PJI2 (Topological Descriptor):** 2D Petitjean Shape Index.
- **IVDE (Information Descriptor):** Mean information content on the vertex degree equality.
- **MATS8v (2D-Autocorrelation Descriptor):** Moran auto correlation –lag 8/weighted by atomic van der waals volumes.
- **MATS2e (2D-Autocorrelation Descriptor):** Moran auto correlation –lag 2/weighted by atomic Sandersons electronegativities.
- **GATS8e (2D-Autocorrelation Descriptor):** Geary auto correlation –lag 8/weighted by atomic Sandersons electronegativities.
- **JGI9 (Topological Charge Descriptor):** Mean topological charge index of order 9.
- **L/Bw (WHIM Descriptor):** Length to breadth ratio by WHIM.
- **R4e+ (GETAWAY Descriptor):** R maximal auto correlation of lag 4/weighted by atomic sanderson electronegativities.
- **nRORPh (Functional Group Counts Descriptor):** Number of ethers.

The reported model has further been used to predict the log K_i values of remaining 20 compounds which are used as test set. The R^2_{pred} comes out to be 0.651 confirming that the proposed model is meaningful. The RMSE and MAE values for the test set are 0.145 and 0.033 respectively.

We notice from the model equation that the most contribution to the binding affinity comes from the two descriptors JGI9 (Mean topological charge index of order 9), it is among the Topological charge indices and MSD (Mean Square Index (Balaban), it is among the topological descriptors. We can conclude that a comparison of results from the model performance demonstrates that the SVM model predicts the binding affinity of the compounds more accurately than ANN and MLR models for the train dataset, while for test set prediction, ANN model was better. Noticed that it might be due to an overfitting problem when a model performs well for calibration but has a poor prediction performance. The proposed models could identify and provide some important information, which is responsible for drug-binding affinity.

The researchers (Sarkar, Middya & Jana, 2012) performed a QSAR study of radical scavenging antioxidant activity of a series of flavonoids using DFT based quantum chemical descriptors. In a search of electronic level understanding of the antioxidant activity of a series of 29 flavonoids with the radical scavenging activity (RSA%) tested in a methanolic solution of DPPH. QSAR studies have been carried out. This is done by using Density Functional Theory (DFT), with the help of Becke's three parameter hybrid density functional, B3LYP/6-31 G(d,p), based quantum chemical descriptors. The best QSAR model have been selected based on their computed Squared Correlation Coefficient of r^2=0.937 and Cross -Validated Squared Correlation Coefficient of q^2 =0.916. The QSAR model indicates that hardness (η), group electrophilic frontier electron density (F^E_A) and group philicity (ω^+_B) of individual molecules are responsible for in vitro biological activity. The group electrophilic frontier electron density (F^E_A) has been used for the first time to explain the radical scavenging activity (RSA) of flavonoids. The excellent correlation between the RSA and the above mentioned DFT based descriptors led to predict new antioxidants (most probably potent compounds). The best MLR model obtained in this study after removing the outlier compounds 11 and 23 was built, and resulted in a r2 and q2 of 0.916 and 0.916, respectively.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Based on the best QSAR model, the authors have designed four new antioxidants, which have a very high radical scavenging ability. So it can be expected that the result of this study should not only help in design of new antioxidants which will prevent free radicals from acting but also facilitate the QSAR study of π-system stabilization in radical form. In conclusion, the excellent QSAR results for flavonoids were obtained using important quantum chemical descriptors which were calculated based on DFT method. From the analysis of the QSAR equation, it can be concluded that in order to get better radical scavenging activity, the hardness, group electrophilic frontier electron density and group electrophilicity should be decreased. In other words, electron-donating capacity and electron-accepting capacity should be reduced. This descriptor explains the contradictory results regarding the antioxidant activities of flavonoids. It depends not on any individual substituents but the effect of the substituents as a whole on the planar ring via the (F^{E}_{A}) descriptor.

The researchers (Rasulev, Abdullaev, Syrov & Leszczynski, 2005) have tried to perform a QSAR study on the antioxidant activity of flavonoids. This QSAR study has been carried out for 27 flavonoids belonging to four different groups (isoflavons, flavons, flavonols, flavanons) to correlate and predict the inhibition of lipid peroxidation effects (antioxidant activity). The Genetic Algorithm (GA) was used to select the descriptors and Multiple Linear Regression analysis (MLRA) was used to generate the correlation models that relate the structural features (descriptors) to the biological activities. The obtained equations consist of one to four descriptors calculated from the characteristics of the molecular structures with use of DRAGON software and quantum-chemical methods. A number of quantum chemical molecular descriptors were obtained from the density functional theory (DFT) B3LYP/6-31G(d, p) level optimized geometries. The results of the GA-MLRA analysis show that the position of the OH groups, the magnitude of dipole moment and the shape of the molecule play an important role in inhibition of lipids peroxidation by flavonoids. The significant QSAR models were obtained with *r*-values of 0.935 and 0.933 for basic models. The q^2 (cross validation r^2) values and scrambling/randomization experiments also confirm the statistical significance of the proposed models. These models are expected to be useful for screening of flavonoid antioxidants. From the best QSAR model, one of the important electronic molecular properties is the dipole moment, which is related to the interaction between a drug and the target molecule. The value of dipole moment defines the orientation of the flavonoid and the interaction rate, which are important in driving the interaction. Data analysis indicates that a good inhibitor of lipid peroxidation should have a small dipole moment (no more then 5 – 6 Debye).

The best QSAR models were built based on μ, I_{Glc} and $I_{OH.}$descriptors, where

- **I_{LPO}:** Inhibition of lipids peroxidation.
- **I_{OH}:** Indicator descriptor, number of OH groups at positions C-3 (C - ring) and the 3',4' position in the B-ring.
- **μ:** Dipole moment.
- **I_{Glc}:** Indicator descriptor, number of Glycoside-like fragments.

In conclusion, this work has shown that a set of specific parameters define the antioxidant activity of flavonoids. One of the main parameters is the indicator variable which embraces the OH groups of the flavonoid core. Other parameters are the quantum-chemical descriptor and the topological descriptor that make it possible to build reliable QSARs. The present analysis of flavonoid QSARs offers insight into their possible mechanisms of action and could be used in the design of new flavonoid-based drugs for the handling of free radical-mediated disease conditions.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

In 2011 (Hemmateenejad, Mehdipour, Deeb, Sanchooli & Miri,) have done a QSAR study related to variable selection in Counter Propagation Neural Networks (CPNN) as a classification tool, with an application to 105 flavonoid compounds of known inhibitory activity against protein tyrosine kinase using substituent electronic descriptors (SED) as novel source of electronic descriptors. In calculating the SED descriptors, the substituents were considered as radical molecules. The molecular structures of all the substituent radical structures were built and full geometry optimization was done for them. Considering no molecular symmetry constraint, all bond lengths and angle optimization were carried out at the level of UHF/6-31G and RHF/6-31G. The calculated descriptors can be classified into three different electronic categories including local charges, dipoles, and orbital energies. The quantum chemical indices of hardness (HD), softness (SOF), electronegativity (EN), and electrophylicity (EPH) were also calculated. Some additional descriptors such as HOMOA (Highest occupied molecular orbital energy), HOMOB, LUMOA (Lowest unoccupied molecular orbital energy), LUMOB, HDB, SOFA, SOFB, ENA, ENB, EPHA, and EPHB stem from two different alpha and beta electronic population energy were calculated where the subscript A and B stand for alpha and beta population of electronic energy, respectively. Therefore, a total of 25 electronic descriptors were calculated for each substituent. Since the studied flavonoid derivatives pose 8 substitution positions, for each molecule a data matrix of SED parameters is obtained by placing the row vectors of different substituents under each other. Consequently, a 3- way data array is obtained by collecting the SED data matrices of different molecules beside each other's.

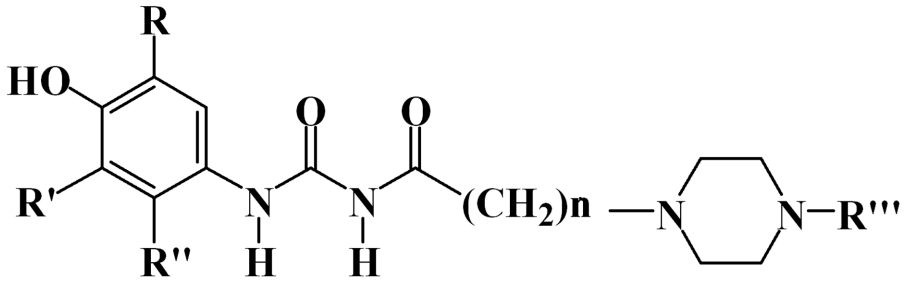
The Classification is a useful tool in pattern recognition of the data set with discrete activity (i.e. active and non-active). Among the classifier methods, Counter Propagation Neural Network has attracted a lot of attention because of its ability to model both discrete and continuous activities. However, feature selection remains a problem. In this study, some filter, wrapper and combinatorial approaches of feature selection and feature extraction were evaluated in order to find optimal procedure for CPNN in a classification problem. Results of all methods showed that F score-based ranking demonstrated the best result while combination of feature extraction method, Principal Component Analysis (PCA) with feature selection process gave poorer results compared to using feature selection alone. In addition, filter methods were better than wrapper methods in this study, since they showed higher performance and also more rational trend in PC-CPNN. Moreover, 3-way handling of data, Parallel Factor Analysis (PARAFAC) gave the worst result which may be due to loss of informative data during transformation process and restricted rotational freedom in calculating the scores. Comparison of CPNN with linear discriminant analysis (LDA), as a routine method of classification in QSAR studies, revealed that applying feature selection to CPNN resulted in better predictions compared to LDA. In conclusion, Counter Propagation Neural Network (CPNN) is one of the most attractive classification tools for QSAR studies. The performance of some different feature selection algorithms was evaluated in order to find the best classification model. The methods were applied for modeling protein-tyrosine kinase inhibitory of 105 flavonoid derivatives using substituent electronic descriptors. Applying feature selection to CPNN resulted in better prediction.

Hydroxyphenylureas

In 2008 (Deeb, Youssef & Hemmateenejad,) performed a QSAR analysis of novel hydroxyphenylureas as antioxidant agents. A set of descriptors consisting of quantum, chemical, and topological descriptors had been used for modeling the activity of 36 novel hydroxyphenylureas to scavenge Free Radicals (FRs) produced by Peripheral Multinuclear Neutrophil Cells (PMNs) expressed as the rate of change per unit added volume of FR (K_{app} min^{-1}). Simple, as well as multiple regressions have indicated that PW2 (Path\

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 2. General structure of hydroxyphenylureas



Walk-2-Randic Shape Index) is the most dominating parameter to be used in modeling K_{app}. Excellent results were obtained in multiparametric regressions. The results of this study are critically discussed using a variety of statistical parameters. The models are validated using Leave-One-Out (LOO) and Leave-Group-Out (LGO) cross validation, external test set, and chance correlation. The general structure of hydroxyphenylureas is given in Figure 2.

The best QSAR model obtained in this study was built using MPC, SSC, S0K, $DM_{z,}$ $Q_{mean,}$ X5, PW2 and DM_t molecular descriptors where:

- **MPC:** Most positive charges,
- **Qmean:** The Mean absolute charge (charge polarization),
- **SSC:** Sum of squares of charges,
- **X5:** Connectivity index Chi-5,
- **S0K:** Kier symmetry index,
- **PW2:** Path/walk-2-randic shape index,
- **DMz:** Molecular dipole moment at z-direction,
- **DM_t:** Total molecular dipole moment.

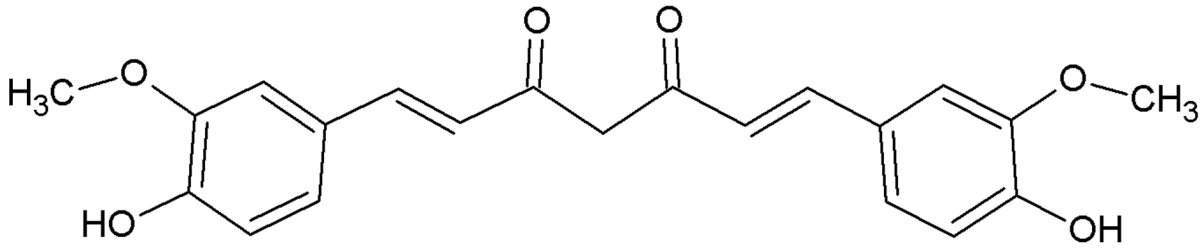
In conclusion, the quantum and topological descriptors can be successfully used for modeling the activity of novel hydroxyphenylureas to scavenge FRs produced by PMNs, expressed as the rate of change per unit added volume of FR Kapp min^{-1}. PW2 is the most dominating parameter to be used in modeling K_{app}.

Curcumin Analogues

In 2014 (Chen, Zhu, Chen, Dong & Li,) performed 3D-QSAR study on the antioxidant capacity or curcumin analogues. A comparative molecular similarity indices analysis (CoMSIA) was done on a set of 27 curcumin-like diarylpentanoid analogues and their DPPH scavenging activities. This study indicates that a combination of steric, hydrophobic, hydrogen bond donor and hydrogen bond acceptor fields, show good correlative and predictive properties. Moreover, the authors have been able to specifically derive chemical properties that are important to activity, and hence adopt a rational approach toward the selection of substituents at various positions in our scaffold; favored and disfavoured regions for enhanced antioxidative activity were suggested. A significant cross-validated correlation coefficient Q^2 = 0.784, SEP = 0.042 for CoMSIA was obtained, indicating the statistical significance of the correla-

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 3. Structure of the curcumin



tion scaffold. The results have been used as a guide to design compounds that, potentially, have better activity against oxidative damage. Figure 3 shows the structure of the curcumin.

Hydroxybenzalacetones

In 2005 (Yamagami, Akamatsu, Motohashi, Hamadaa & Tanahashi,) performed a QSAR study for antioxidant hydroxybenzalacetones by quantum chemical and 3-D-QSAR (CoMFA) analysis. Antioxidant activities for a series of hydroxybenzalacetones, OH-BZ, evaluated by their inhibitory potencies against lipid peroxidation induced by c-ray irradiation or t-BuOOH, were analyzed quantitatively using quantum-chemical parameters calculated by semi-empirical molecular orbital calculations. The energy of the highest occupied molecular orbital (E_{HOMO}) and frontier electron densities on the phenolic oxygen atom ($F_{H,O}$), together with the steric parameter (E_s) for the substituent ortho to the phenolic oxygen, showed excellent correlations. The classical QSAR approach and the approach in this study led the authors to conclude that the activity is increased by electron-donating substituents and/or by bulky ortho substituents: the former is likely to facilitate the hydrogen abstraction from phenols and the latter is considered to function to stabilize the resultant phenoxy radical. Although classical QSAR using various experimentally derived physicochemical parameters affords more straightforward information about the reaction mechanism, the successful prediction of activity by quantum chemical QSARs in terms of parameters obtained by rapid semiempirical molecular orbital calculations will permit their wider application in the search for antioxidant compounds. The authors also performed 3D-QSAR studies by using the comparative molecular field analysis (CoMFA) model.

In conclusion, the best QSAR model was obtained in this study by using quantum chemical descriptors such as HOMO, ΣF H,O and ΣEs as well as indicator variable (Ip).

The definition of these variables in the QSAR model is given in the above text. This result demonstrates that the higher the HOMO energy and frontier electron density (HOMO) on phenolic O-atom(s), and the bulkier ortho substituent(s), the greater the antioxidant activity is. Since the former two factors are related to the facility of electrondonating ability of substituents, the QSARs obtained here are considered to have a physical meaning comparable to that of the previous classical QSARs.

Phenolic Derivatives Bearing No Donor Groups

In 2011 (Mitra, Saha & Roy,) performed a QSAR study of 33 antioxidant compounds belonging to the class of phenolic derivatives bearing NO donor groups. This dataset exhibiting a wide range of antioxidant activity (IC_{50}). The antioxidant activities of the compounds were measured using the TBARS

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

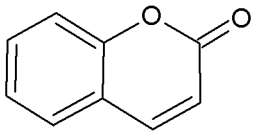
(Thiobarbituric acid reactive substance) assay method. Different types of descriptors were calculated in this study such as topological indices, structural, thermodynamic, spatial and electronic descriptors. The chemometric tools used in this study were the Genetic Function Approximation (GFA) and Genetic Partial Least Squares (G/PLS) techniques. Initially, QSAR model was developed using the entire data set using stepwise regression. In order to determine the external predictive ability of models, the data set was divided into training (25 compounds) and test (8 compounds) sets using the *k*-means clustering technique and external validation was done based on the activity prediction of the test set compounds. $R^2_{pred} = 0.917$ (for the test set) indicates significant ability of the developed model to predict the activity of new compounds belonging to this series of phenolic derivatives. A randomization test was performed, and the results indicate that the developed models are sufficiently robust and not the outcome of mere chance. In conclusion, the present work deals with QSAR studies of a series of antioxidants belonging to the class of phenolic derivatives bearing NO donor groups. Several QSAR models with appreciable statistical significance have been reported. Models were built using various chemometric tools, which were validated both internally and externally. These models chiefly infer that presence of substituted aromatic carbons, long chain branched substituents, an oxadiazole-N-oxide ring with an electronegative atom containing group and high degree of methyl substitutions of the parent moiety are conducive to the antioxidant activity profile of these molecules. The novelty of this study is not only that the structural attributes of NO donor phenolic compounds required for potent antioxidant activity have been explored, but 15 new compounds with possible antioxidant activity have also been designed and their antioxidant activity has been predicted *in silico*.

Coumarin Derivatives

Recently (Mitra, Saha & Roy, 2013) performed predictive modeling of antioxidant coumarin derivatives using different approaches that include: Descriptor based QSAR, 3 dimensional pharmacophore mapping and Hologram Quantitative Structure-Activity Relationship (HQSAR). The study deals with a series of coumarin derivatives (45 compounds) that were modeled for their antioxidant activity based on their ability to inhibit DPPH (1,1-diphenyl-2-picryl-hyrazyl) free radicals. The descriptor-based QSAR model provides quantitative insight about the different molecular pre-requisites for exhibiting potential antioxidant activity. The three dimensional pharmacophore model and the HQSAR model enable the detection of the different molecular features and the atomic fragments that are important to the antioxidant activity of the coumarin derivatives. Figure 4 shows the parent coumarin nucleus used in this study:

The three different types of QSAR models developed in this study simplify the process of identifying the essential molecular fragments and in that way facilitate the selection of molecules exhibiting improved radical scavenging activity. The descriptor-based QSAR model was developed using two different chemometric tools which include the Genetic Function Approximation (GFA) and Genetic Partial Least Squares (G/PLS) techniques based on both linear and spline options. Based on the differ-

Figure 4. Parent coumarin nucleus used in this study



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

ent internal and external validation metrics, the G/PLS model developed using the spline option was selected as the best one. The model infers that compounds bearing a higher degree of branching, which then leads to a decrease in their molecular volume, exhibit enhanced free radical scavenging activity. In addition, molecules with an oxygen atom-bearing fragment (-OH, phenolic fragment, carbonyl fragment =O and others) as well as a tertiary carbon as substituents also lie in the higher activity range. For the three-dimensional pharmacophore model that was developed using the *BEST* method of conformer generation, the pharmacophoric features responsible for the explicit biological activity of the molecules was identified. The best pharmacophore hypothesis, selected based on the Correlation Coefficient and the cost functions of the developed models, indicate that three hydrogen bond acceptor features and a hydrophobic feature lying at the specified distances from each other were chiefly responsible for the optimum antioxidant activity profile of the molecules. The importance of the hydrogen bond acceptor features was in good agreement with the descriptor-based QSAR model, which reveals the significance of oxygen atom-bearing fragments to the overall activity profile of the molecules. Furthermore, the contribution of the different molecular fragments to the overall activity of the coumarin derivatives was identified based on the HQSAR analysis. The contribution map for the HQSAR model revealed the importance of the parent coumarin nucleus for the optimal activity of the molecules. The results appropriately go with those of the pharmacophore analysis, which indicates the importance of the fused benzene ring and the oxygen atom of the pyran ring for capturing the hydrophobic feature and one of the hydrogen bond acceptor features, respectively. Consequently, the diverse QSAR techniques applied in this study correlate well with each other. As a conclusion, this work outlines the importance of the different molecular fragments or features at various positions favoring the antioxidant activity profile as inferred from the three different models: descriptor based QSAR, 3D pharmacophore and HQSAR model.

In 2012 (Martínez-Martínez, Razo-Hernández, Peraza-Campos, Villanueva-García, Sumaya-Martínez, Cano & Gómez-Sandoval,) performed synthesis and *in vitro* antioxidant activity evaluation of 3-carboxycoumarin derivatives and QSAR study of their DPPH radical scavenging activity. In this study, the *in vitro* antioxidant activities of eight 3-carboxycoumarin derivatives were assayed by the quantitative 1,1-diphenyl-2-picrylhydrazil DPPH radical scavenging activity method. The two compounds 3-Acetyl-6-hydroxy-2*H*-1-benzopyran-2-one and ethyl 6-hydroxy-2-oxo-2*H*-1-benzopyran-3-carboxylate presented the best radical-scavenging activity. A QSAR study was performed and correlated with the experimental DPPH scavenging data. The authors used structural, geometrical, topological and quantum-chemical descriptors selected with Genetic Algorithms in order to determine which of these parameters are responsible of the observed DPPH radical scavenging activity. The authors developed a Back Propagation Neural Network (BPNN) model with the hydrophilic factor (Hy) descriptor to generate an adequate architecture of neurons for the system description. The mathematical model showed a multiple determination coefficient of 0.9196 and a Root Mean Squared Error of 0.0851. In genetic analysis, the average percent error obtained was 3.77% while in back propagation neutral network, the average percent error was 7.18%. This result indicates that the combination of the two methodologies optimize the creation of QSAR models. The genetic analysis allows finding the most important descriptor for the development of the antiradical activity and artificial neural network improves the model with the use of only one molecular descriptor to obtain accurate prediction values. The presence of hydroxyl groups on the ring structure of 3-carboxycoumarins is correlated with their DPPH radical scavenging effects. The mixed QSAR model showed that hydrophilic factor (Hy) descriptor could indicate that antiradical activity would increase as we incorporate hydroxyl groups in the coumarin molecules. In conclusion, this study indicates the importance of the –OH hydrophilic group as crucial for antiradical activity of coumarins.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

In other words, the models indicate that the antiradical activity increases as we incorporate hydrophilic groups to the coumarin molecules.

CONCLUSION

Computational-based rational drug design has become more common over the past decade. Most of the approaches involved in this type of design focus on Quantitative Structure–Activity Relationship (QSAR) studies that use various molecular descriptors for chemical information training. QSARs represent an attempt to correlate physicochemical or structural descriptors of a set of compounds with their biological (pharmacological, toxicological) activities.

In this chapter, we have summarized all effort done in building QSAR models of antioxidants. The activity of the antioxidant compounds as well as their structural variety makes them a rich resource for modeling lead compounds with targeted pharmacological properties. Different QSAR studies have been reviewed related to flavonoids, hydroxyphenylureas, curcumin analogues, hydroxybenzalacetones, phenolic derivatives bearing NO donor groups and coumarin derivatives. Therefore, the outcome of this study could be used as a guide for further development of selective or the predictive ability of newly designed antioxidative analogues, prior to synthesis.

In conclusion, different *in silico* methods were presented in this chapter especially QSAR in order to account for the activity of different antioxidant compounds. In these QSAR studies, different statistical methods were used in order to build QSAR models using different types of descriptors. As it can be seen from these studies, the authors tried to use the best validated QSAR models in order to design new compounds with increased antioxidant activity. Therefore, it seems that designing new antioxidant compounds using QSAR methodology requires novel descriptors such as the work done by Deeb and Clare (Deeb & Clare, 2007) in which they have investigated the flavonoid-inhibitory activity of 54 analogues using the nodal angle descriptors (Clare 2000; Clare & Supuran, 2005) and flipstep regression analysis. Also a combination of QSAR and docking or other methods will be the choice in designing new antioxidant compounds. For example, the recent study (Mitra, Saha & Roy, 2013) in which the authors performed predictive modeling of antioxidant coumarin derivatives using different approaches that include: Descriptor based QSAR, 3 dimensional pharmacophore mapping and Hologram Quantitative Structure-Activity Relationship (HQSAR).

Many QSAR studies have been performed regarding the antioxidant activity of different series of compounds. However, these QSAR studies are similar and sometimes repeated on the same class of compounds with a minor improvement. It can be suggested that the use of innovative or novel descriptors in the QSAR studies could be a new direction for this field of study. The suggested method will assure the designing of new compounds with antioxidant activity.

REFERENCES

Allen, D. M. (1974). The relationship between variable selection and data augmentation and a method for prediction. *Technometric*, *16*(1), 125–127. doi:10.1080/00401706.1974.10489157

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Antolovic, M., Prenzler, P. D., Patsalides, E., McDonald, S., & Robards, K. (2002). Methods for testing antioxidant activity. *Analyst (London)*, *127*(1), 183–198. doi:10.1039/b009171p PMID:11827390

Benzie, I. F. (2003). Evolution of dietary antioxidants. *Comparative Biochemistry and Physiology. Part A, Molecular & Integrative Physiology*, *136*(1), 113–126. doi:10.1016/S1095-6433(02)00368-9 PMID:14527634

Chen, B., Zhu, Z., Chen, M., Dong, W., & Li, Z. (2014). Three-dimensional quantitative structure–activity relationship study on antioxidant capacity of curcumin analogues. *Journal of Molecular Structure*, *1061*, 134–139. doi:10.1016/j.molstruc.2013.12.083

Clare, B. W. (1998). The frontier orbital phase angles: Novel QSAR descriptors for benzene derivatives, applied to phenylalkylamine hallucinogens. *Journal of Medicinal Chemistry*, *41*(20), 3845–3856. doi:10.1021/jm980144c PMID:9748359

Clare, B. W. (2000). The frontier orbital phase angles: A theoretical interpretation. *Theochem*, *507*(1-3), 157–164. doi:10.1016/S0166-1280(99)00392-9

Clare, B. W. (2001a). Erratum to "The frontier orbital phase angles: A theoretical interpretation". *Theochem*, *535*(1-3), 301–301. doi:10.1016/S0166-1280(00)00847-2

Clare, B. W., & Supuran, C. T. (2005). A physically interpretable quantum-theoretic QSAR for some carbonic anhydrase inhibitors with diverse aromatic rings, obtained by a new QSAR procedure. *Bioorganic & Medicinal Chemistry*, *13*(6), 2197–2211. doi:10.1016/j.bmc.2004.12.055 PMID:15727872

Clare, B. W., & Supuran, C. T. (2005a). Predictive flip regression: A technique for QSAR of derivatives of symmetric molecules. *Journal of Chemical Information and Modeling*, *45*(5), 1385–1391. doi:10.1021/ci050191v PMID:16180915

Deeb, O., & Clare, B. W. (2007). QSAR of aromatic substances: Protein tyrosine kinase inhibitory activity of flavonoid analogues. *Chemical Biology & Drug Design*, *70*(5), 437–449. doi:10.1111/j.1747-0285.2007.00578.x PMID:17927721

Deeb, O., Shaik, B., & Agrawal, V. (2014). Exploring QSARs of the interaction of flavonoids with GABA (A) receptor using MLR, ANN and SVM techniques. *Journal of Enzyme Inhibition and Medicinal Chemistry*, *29*(5), 670–676. doi:10.3109/14756366.2013.839557 PMID:24102524

Deeb, O., Youssef, K., & Hemmateenejad, B. (2008). QSAR of novel hydroxyphenylureas as antioxidant agents. *QSAR & Combinatorial Science*, *27*(4), 417–424. doi:10.1002/qsar.200730023

Devillers, J., & Balaban, A. T. (1999). *Topological indices and related descriptors in QSAR and QSPR*. Amsterdam: Gordon Breach Scientific Publishers.

Downs, G. M. (2004). Molecular descriptors. In Computational medicinal chemistry for drug discovery (pp. 515-538). Marcel Dekker.

Dragan, A., Dusanka, D., Drago, B., Vesna, R., Bono, L., & Nenad, T. (2007). Article. *Current Medicinal Chemistry*, *14*, 827–845. doi:10.2174/092986707780090954 PMID:17346166

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Duprat, A. F., Huynh, T., & Dreyfus, G. (1998). Toward a principled methodology for neural network design and performance evaluation in QSAR: Application to the prediction of log P. *Journal of Chemical Information and Computer Sciences*, *38*(4), 586–594. doi:10.1021/ci980042v PMID:9691473

Eriksson, L., Johansson, E., Kettaneh-Wold, N., & Wold, S. (2001). *Multi- and megavariate data analysis. principles and applications*. Umea: Umetrics.

Free, S. M., & Wilson, J. W. (1964). A mathematical contribution to structure-activity studies. *Journal of Medicinal Chemistry*, *7*(4), 395–399. doi:10.1021/jm00334a001 PMID:14221113

Fujita, T. (1990). The extrathermodynamic approach to drug design. In Comprehensive medicinal chemistry (Vol. 4). Pergamon.

Fujita, T., Iwasa, J., & Hansch, C. (1964a). A new substituent constant, π, derived from partition coefficients. *Journal of the American Chemical Society*, *86*(23), 5175–5180. doi:10.1021/ja01077a028

Gunn, S. R. (1997). *Support vector machines for classification and regression*. University of Southampton.

Halvorsen, B. L., Carlsen, M. H., Phillips, K. M., Bohn, S. K., & Holte, K. (2006). Content of redox-active compounds (i.e., antioxidants) in foods consumed in the United States. *The American Journal of Clinical Nutrition*, *84*, 95–135. PMID:16825686

Halvorsen, B. L., Holte, K., Myhrstad, M. C. W., Barikmo, I., & Hvattum, E. (2002). A systematic screening of total antioxidants in dietary plants. *The Journal of Nutrition*, *132*, 461–471. PMID:11880572

Hansch, C. (1969). A quantitative approach to biochemical structure-activity relationships. *Accounts of Chemical Research*, *2*(8), 232–239. doi:10.1021/ar50020a002

Hansch, C., & Fujita, T. (1964b). ρ−σ−π analysis. A method for the correlation of biological activity and chemical structure. *Journal of the American Chemical Society*, *86*(8), 1616–1626. doi:10.1021/ja01062a035

Hansch, C., Maloney, P., Fujita, T., & Muir, R. M. (1962). Correlation of biological activity of phenoxyacetic acids with hammett substituent constants and partition coefficients. *Nature*, *194*(4824), 178–180. doi:10.1038/194178b0

Hemmateenejad, B., Mehdipour, A., Deeb, O., Sanchooli, M., & Miri, R. (2011). Toward an optimal approach for variable selection in counter-propagation neural networks: Modeling protein-tyrosine kinase inhibitory of flavanoids using substituent electronic descriptors. *Molecular Informatics.*, *30*(11-12), 939–949. doi:10.1002/minf.201100081

Jayaprakasha, G. K., Jena, B. S., Negi, P. S., & Sakariah, K. K. (2002). Evaluation of antioxidant activities and antimutagenicity of turmeric oil: A byproduct from curcumin production. *Zeitschrift fur Naturforschung.*, *57c*, 828–835. PMID:12440720

Jolliffe, I. T. (1986). *Principal component analysis*. New York: Springer-Verlag. doi:10.1007/978-1-4757-1904-8

Karelson, M. (2000). *Molecular descriptors in QSAR/QSPR*. New York: Wiley-InterScience.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Katritzky, A. R., Lobanov, V. S., & Karelson, M. (1994). CODESSA, reference manual. University of Florida. Retrieved from http://www.semichem.com/codessa

Knight, J. (1998). Free radicals: Their history and current status in aging and disease. *Annals of Clinical and Laboratory Science*, *28*(6), 331–346. PMID:9846200

Martínez-Martínez, J., Razo-Hernández, R., Peraza-Campos, A., Villanueva-García, M., Sumaya-Martínez, M., Cano, D., & Gómez-Sandoval, Z. (2012). Synthesis and *in vitro* antioxidant activity evaluation of 3-carboxycoumarin derivatives and QSAR study of their DPPH radical scavenging activity. *Molecules (Basel, Switzerland)*, *17*(12), 14882–14898. doi:10.3390/molecules171214882 PMID:23519260

Matill, H. A. (1947). Antioxidants. *Annual Review of Biochemistry*, *16*(1), 177–192. doi:10.1146/annurev.bi.16.070147.001141 PMID:20259061

Milardovic, S., Ivekovic, D., & Grabaric, B. S. (2006). A novel amperometric method for antioxidant activity determination using DPPH free radical. *Bioelectrochemistry (Amsterdam, Netherlands)*, *68*(2), 175–180. doi:10.1016/j.bioelechem.2005.06.005 PMID:16139574

Mitra, I., Saha, A., & Roy, K. (2011). Chemometric QSAR modeling and in silico design of antioxidant no donor phenols. *Scientia Pharmaceutica*, *79*(1), 31–57. doi:10.3797/scipharm.1011-02 PMID:21617771

Mitra, I., Saha, A., & Roy, K. (2013). Predictive modeling of antioxidant coumarin derivatives using multiple approaches: descriptor-based QSAR, 3D-pharmacophore mapping, and HQSAR. *Scientia Pharmaceutica*, *81*(1), 57–80. doi:10.3797/scipharm.1208-01 PMID:23641329

Montgomery, D. C., & Peck, E. A. (1992). *Introduction to linear regression analysis*. New York: Wiley.

Novi, M., Nikolovska-Coleska, Z., & Solmajer, T. (1997). Quantitative structure-activity relationship of flavonoid p56 protein tyrosine kinase inhibitors. A neural network approach. *Journal of Chemical Information and Computer Sciences*, *37*(6), 990–998. doi:10.1021/ci970222p

Pisoschi, A. M., Cheregi, M. C., & Danet, A. F. (2009). Total antioxidant capacity of some commercial fruit juices: Electrochemical and spectrophotometrical approaches. *Molecules (Basel, Switzerland)*, *14*(1), 480–493. doi:10.3390/molecules14010480 PMID:19158657

Rasulev, B., Abdullaev, N., Syrov, V., & Leszczynski, J. (2005). A quantitative structure-activity relationship (QSAR) study of the antioxidant activity of flavonoids. *QSAR & Combinatorial Science*, *24*(9), 1056–1065. doi:10.1002/qsar.200430013

Sarkar, A., Middya, T. R., & Jana, A. D. (2012). A QSAR study of radical scavenging antioxidant activity of a series of flavonoids using DFT based quantum chemical descriptors – The importance of group frontier electron density. *Journal of Molecular Modeling*, *18*(6), 2621–2631. doi:10.1007/s00894-011-1274-2 PMID:22080306

Sies, H. (1997). Oxidative stress: Oxidants and antioxidants. *Experimental Physiology*, *82*(2), 291–295. doi:10.1113/expphysiol.1997.sp004024 PMID:9129943

Silipo, C., & Vittoria, A. (1990). Three-dimensional structure of drugs. *Medicinal Chemistry (Shariqah, United Arab Emirates)*, *4*, 154–204.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Tetko, I. V., Alessandro, E., Villa, P., & Livingstone, D. J. (1996). Neural network studies. 2. Variable selection. *Journal of Chemical Information and Computer Sciences*, *36*(4), 794–803. doi:10.1021/ci950204c PMID:8768768

Thaipong, K., Boonprakob, U., Crosby, K., Cisneros-Zevallos, L., & Byrne, D. H. (2006). Comparison of ABTS, DPPH, FRAP, and ORAC assays for estimating antioxidant activity from guava fruit extracts. *Journal of Food Composition and Analysis*, *19*(6-7), 669–675. doi:10.1016/j.jfca.2006.01.003

Todeschini, R., Consonni, V., & Pavan, M. (2001). *DRAGON-software for the calculation of molecular descriptors. Release 1.12 for Windows*. Retrieved from http://www.disat.unimib.it/chm

Valko, M., Leibfritz, D., Moncol, J., Cronin, M., Mazur, M., & Telser, J. (2007). Free radicals and antioxidants in normal physiological functions and human disease. *The International Journal of Biochemistry & Cell Biology*, *39*(1), 44–84. doi:10.1016/j.biocel.2006.07.001 PMID:16978905

Vertuani, S., Angusti, A., & Manfredini, S. (2004). The antioxidants and proantioxidants network: An overview. *Current Pharmaceutical Design*, *10*(14), 1677–1694. doi:10.2174/1381612043384655 PMID:15134565

Wold, S. (1978a). Cross-validatory estimation of the number of components in factor and principal component models. *Technometrics*, *20*(4), 397–405. doi:10.1080/00401706.1978.10489693

Wold, S. (1991b). Validation of QSARs. *Quantitative Structure Activity Relationship*, *10*(3), 191–193. doi:10.1002/qsar.19910100302

Wold, S., Johansson, E., & Cocchi, M. (1993). PLS—Partial least squares projections to latent structures. In 3D-QSAR in drug design, theory, methods, and applications (pp. 523–550). ESCOM Science Publishers.

Wolf, G. (2005). The discovery of the antioxidant function of vitamin E: The contribution of Henry A. Mattill. *The Journal of Nutrition*, *135*(3), 363–366. PMID:15735064

Xu, L., & Zhang, W. J. (2001). Comparison of different methods for variable selection. *Analytica Chimica Acta*, *446*(1-2), 477–483. doi:10.1016/S0003-2670(01)01271-5

Yamagami, C., Akamatsu, M., Motohashi, N., Hamadaa, S., & Tanahashia, T. (2005). Quantitative structure–activity relationship studies for antioxidant hydroxybenzalacetones by quantum chemical and 3-D-QSAR (CoMFA) analysis. *Bioorganic & Medicinal Chemistry Letters*, *15*(11), 2845–2850. doi:10.1016/j.bmcl.2005.03.087 PMID:15911266

KEY TERMS AND DEFINITIONS

Antioxidant: A stable molecule that donates an electron to a free radical and neutralizes it, therefore reducing its capability to damage.

Computer Aided-Drug Design (CADD): The effective design of chemical structures with the desirable therapeutic properties, it is a well-established area of computer aided molecular design (CAMD).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Curcumin: Naturally phenolic compound which is isolated from Curcuma longa Linn. It is widely used as food pigment and it has been reported to have many biological activities especially the favorable effect on antioxidation.

Flavonoids: Groups of naturally occurring compounds of low molecular weight plant products, based on the parent compound, flavone (2-phenylchromone) and have shown potential for application in a variety of pharmacological targets.

Free Radicals: Unstable molecules that lose one of its electrons and therefore become unbalanced and highly reactive.

In Silico Methods: New developments in chemical testing that rely on computer simulation or modeling. *In silico* pharmacology defines the use of this information in the design of computational models or simulations that can be used to make predictions, suggest hypotheses, and ultimately supply discoveries in therapeutics.

QSAR: Quantitative structure activity relationships is a way of finding a simple equation that correlates structural molecular features (descriptors) with physicochemical properties, such as biological activities for a set of compounds by means of statistical methods.

This work was previously published in Quantitative Structure-Activity Relationships in Drug Design, Predictive Toxicology, and Risk Assessment edited by Kunal Roy, pages 212-237, copyright year 2015 by Medical Information Science Reference (an imprint of IGI Global).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 16
Integrated *in Silico* Methods for the Design and Optimization of Novel Drug Candidates: A Case Study on Fluoroquinolones – *Mycobacterium tuberculosis* DNA Gyrase Inhibitors

Nikola Minovski
National Institute of Chemistry, Slovenia

Marjana Novič
National Institute of Chemistry, Slovenia

ABSTRACT

Although almost fully automated, the discovery of novel, effective, and safe drugs is still a long-term and highly expensive process. Consequently, the need for fleet, rational, and cost-efficient development of novel drugs is crucial, and nowadays the advanced in silico drug design methodologies seem to effectively meet these issues. The aim of this chapter is to provide a comprehensive overview of some of the current trends and advances in the in silico design of novel drug candidates with a special emphasis on 6-fluoroquinolone (6-FQ) antibacterials as potential novel Mycobacterium tuberculosis DNA gyrase inhibitors. In particular, the chapter covers some of the recent aspects of a wide range of in silico drug discovery approaches including multidimensional machine-learning methods, ligand-based and structure-based methodologies, as well as their proficient combination and integration into an intelligent virtual screening protocol for design and optimization of novel 6-FQ analogs.

DOI: 10.4018/978-1-5225-0549-5.ch016

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

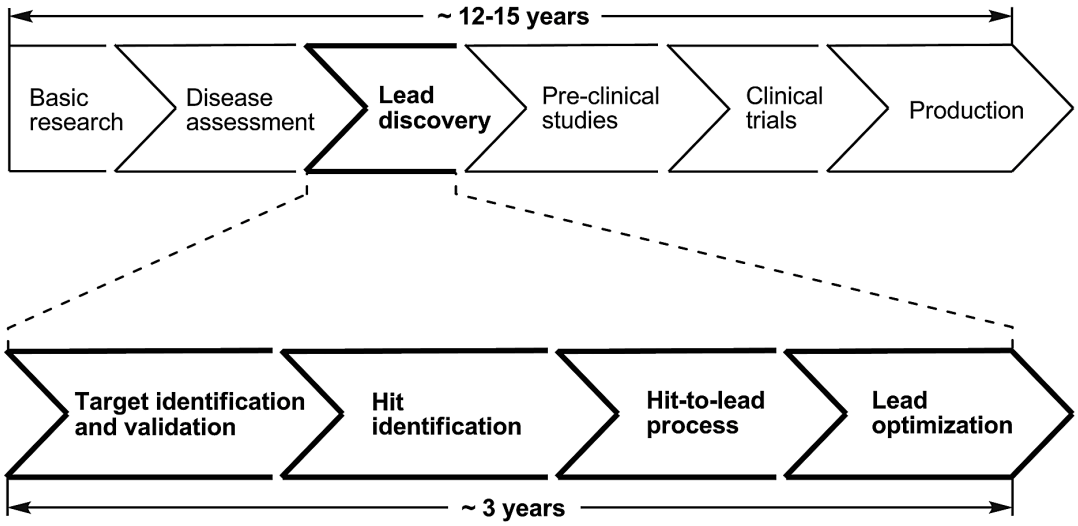
INTRODUCTION

The discovery process of novel, effective and safe drugs, from day to day is becoming more advanced and sophisticated. Today, the profit- and innovation-based competition between the pharmaceutical companies is increasingly growing (Cavazzani, 2010). In addition to the profit, the process of discovery a new drug is not only long-term, but also highly expensive. It was estimated that a new drug discovery program is approximately 12-15 years long and takes around 200-300 millions to one billion dollars (Rawlins, 2004; Bartfai & Lees, 2006; Hughes et al., 2011). To alleviate this growing problem, nowadays the drug research efforts are mainly directed toward reducing the discovery costs as well as the time required. In that regard, the *in silico* drug discovery methods were found as particularly important to effectively meet these issues.

As depicted in Figure 1, the lead discovery part (which is constructed of several interconnected subphases) of the drug discovery pipeline can be considered as the essential one (Kenny et al., 1998; Langer & Hoffmann, 2001). Owing to the rapid development of various computational methods that today could readily and efficiently be applied in different lead discovery segments, one can observe a significant progress in reducing its durability to amazing 2-3 years (Figure 1). However, within the framework of the entire drug discovery process it is still a long period of time and therefore a further time reduction is certainly welcomed.

In the seventies of the previous century, the discovery of novel lead compounds was substantially based on a random screening of large chemical libraries comprised of chemicals of different origin. This approach has been initially used for discovery of new antibiotics. The drug discovery practice demonstrated that on average only one potential lead from a library containing around 20.000 molecules could be identified using the random screening approach (Young et al., 1996). Since the 1980s, with the growth and development of robotics and miniaturization of the *in vitro* testing methods, it became possible to screen hundreds of thousands of compounds on a large number of biological targets (Gribbon & Sewing, 2005), i.e., a methodology widely known as high-throughput screening (HTS). Nevertheless, such

Figure 1. Schematic representation of a classical industrial drug discovery pipeline



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

a philosophy postulated under the idea, the greater is the starting chemical library, the higher are the chances to identify a biologically-active molecule, was disappointing for many pharmaceutical companies (O'Driscoll, 2004). These failures as well as the daily progresses in molecular and structural biology were the major driving force to elevate the discovery of novel drugs to a significantly higher, knowledge-based level commonly known as the rational drug design (Mavromoustakos et al., 2011). Moreover, the advances in the computation and strategies such as 2D/3D computer-aided molecular design (CAMD) opened a new perspective into the drug discovery world. Nowadays, the *in silico* screening methods are indeed an efficient supplement to the experimentally grounded HTS methods, becoming an integral segment of the hit identification and lead generation processes (Klebe, 2006; Stahura & Bajorath, 2004; Bajorath, 2002; Shoichet, 2004; Bleicher et al., 2003).

The present text aim to report some of the recent trends and advances in the *in silico* design of novel drug candidates – from chemical sketches to predicted active conformations. Specifically organized in two, tutorial-like parts (theoretical and practical), this chapter is not strictly intended to expose the thorough theoretical and mathematical background behind the *in silico* methodologies employed, but rather to guide the reader through the different steps of the *in silico* design and prediction of novel biologically-active hits. Put differently, the theoretical part gives an overview of various attentively selected computational methodologies proficiently integrated into an efficient *in silico* screening framework including quantitative structure-activity relationship (QSAR) methods, virtual combinatorial library design, property-based screening and construction of focused "drug-like" libraries as well as ligand- and structure-based methodologies coupled with virtual screening (VS) for identification and selection of novel hit compounds. On the other hand, the second part of this chapter is primarily focused on the practical implementation of these methods for design of 6-fluoroquinolone (6-FQ) antibacterials as potential novel *Mycobacterium tuberculosis* DNA gyrase inhibitors. It covers all the aspects described in the theoretical part, but proficiently assembled and integrated into an intelligent VS protocol that could aid not only the development of new drugs in general, but also to shorten the time required.

Background

During the last five decades, a wide variety of *in silico* drug design approaches have been invented that undoubtedly altered the drug discovery paradigm – a rapid and cost-efficient development of potent and safe drugs. Nowadays, we have witnessed the effectiveness and benefits of these methods as some of them have not only facilitated the drug discovery in its entirety, but also became a golden standard to success (Ekins et al., 2007; Bharath et al., 2011; Cumming et al., 2013).

From the plethora of computational methods available today, just a few can be regarded as of exceptional importance for the purposes of modern drug discovery. From simple numerical predictions of biological activity values for compounds, which have not yet been synthesized, to possible intuitively based revelation of compound-protein interactions, these methods could be recognized as pivotal indivisible approaches within the framework of the entire drug discovery process. This list of *in silico* drug design approaches includes machine-learning (statistical) methods such as the QSAR, the methodology for virtual combinatorial library design, the property-based filtering approaches as well as the ligand- and structure-based methodologies (Ekins et al., 2007). Notwithstanding their confirmed high potential in various lead discovery segments, one cannot expect a successful identification of novel drug candidates when these methods are used separately. Consequently, their proficient assemblage into powerful *in silico* screening protocols for fast and accurate predictions of novel compounds can be regarded as

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

a major challenge of many computational chemists (Salemme et al., 1997; Lewis, 2005; Nevin et al., 2012; Tian et al., 2013; Chen, 2013).

In the quest for more sophisticated and efficient ways toward the computational design of novel drug candidates, various integrated *in silico* drug design methodologies are developed and a bundle of useful examples can be found in the literature. A straightforward and frequently utilized *in silico* integrations is the methodology of extrapolation of a pre-devised QSAR model for selection and/or prediction of the biological activity values for compounds comprising a virtual combinatorial library (Liu et al., 1998; Burden & Winkler, 1999; Agrafiotis, 2000; Andersson et al., 2001; Grzybowski et al., 2002; Cruz-Monteagudo et al., 2008). As demonstrated, principal component analysis (PCA), artificial neural networks (ANNs) or 3D-QSAR models such as the Comparative Molecular Field Analysis (CoMFA) are the most widely exploited as robust *in silico* screening devices for selection of hit candidates, while genetic algorithm (GA) or desirability-based methods (e.g., implementation of multi-objective optimization for simultaneous mapping of desired drug properties within QSAR models) are commonly applied for ranking purposes (Cruz-Monteagudo et al., 2008; Nicolaou & Brown, 2013). Notwithstanding their favorable outcome in many cases (Grzybowski et al., 2002), it must be taken into account that not every QSAR model is suitable. The lack of validation of the established QSAR model, the wrongly assessed or not assessed model's applicability domain, or even the use of mis- and/or non-interpretable molecular descriptors are just a few instances which could lead the drug discovery process into a wrong direction (Scior et al., 2009; Bajot, 2010; Sacan et al., 2012). Even in cases where the QSAR model is properly established, its implementation just as a rough *in silico* filtering tool for selection of drug candidates from a massive pool of possibilities without taking into consideration their druggability properties (Oprea, 2000) could also outcome in disappointingly low hit rates. Moreover, the absence of any structural data for the biological target or even an *a priori* knowledge for the structure-activity relationship (SAR) between the investigated entities, could additionally contribute to a poor epilogue of the screening strategy.

Conceptually similar, but improved strategies that cover all of these missing aspects or just partially, are the methods of employing three-dimensional (3D) structural concepts such as the ligand-based, structure-based, or their combination as *in silico* filters for virtual combinatorial libraries (Hecker et al., 2002; Salemme et al., 1997; Shaikh et al., 2007; Krovat et al., 2005; Schlegel et al., 2007; Vilar et al., 2009; Zhang et al., 2013; Drwal & Griffith, 2013). As implied by their name, the availability of 3D structural information for the ligand(s) and/or biological target(s) usually obtained by X-ray crystallography or NMR is of critical significance for their practical utilization as *in silico* ligand filters. More importantly, these methods could also give a structural insight into the possible ligand-protein interactions – an invaluable information that indicate to, what might be expected or what will be the next step toward the design of better hits. Over the years, various straightforward, but also advanced approaches have been invented and some of them became a cornerstone of the modern drug discovery. Among them, the ligand-based and its more advanced counterpart the structure-based pharmacophore modeling concept as well as the methodology of molecular docking of ligands into a defined protein binding site are the most commonly utilized for virtual ligand screening (Wolber & Langer, 2005; Wolber & Kosara, 2006; Schlegel et al., 2007).

However, the majority of these methods are principally not designed for *in silico* screening of extremely massive compound libraries (small-scale screening methods), but rather reduced target-focused libraries (Harris et al., 2011), which can be obtained by integration of various rational knowledge-based pre-processing strategies (e.g., QSAR-based activity predictions and screening, "drug-likeness" filtering,

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 1. A summary of various successful in silico drug design integrations (LBD, ligand-based design; SBD, structure-based design)

***In Silico* Drug Design Methodologies**				**References**
QSAR	**Drug-Likeness Assessment**	**LBD**	**SBD**	
+	+	+	-	Khalaf et al., 2010; & Abuhamdah et al., 2013
+	-	+	-	Yang et al., 2013 ; & Kamaria et al., 2014
+	+/-	-	+	Hu et al., 2013; Tan et al., 2013; Ul-Haq et al., 2013; & Zheng et al., 2014
+	+/-	+	+	Musmuca et al., 2010; Coi et al., 2013; Gozalbes et al., 2013
+	+	+	+	Shaikh et al., 2007; Tian et al., 2013

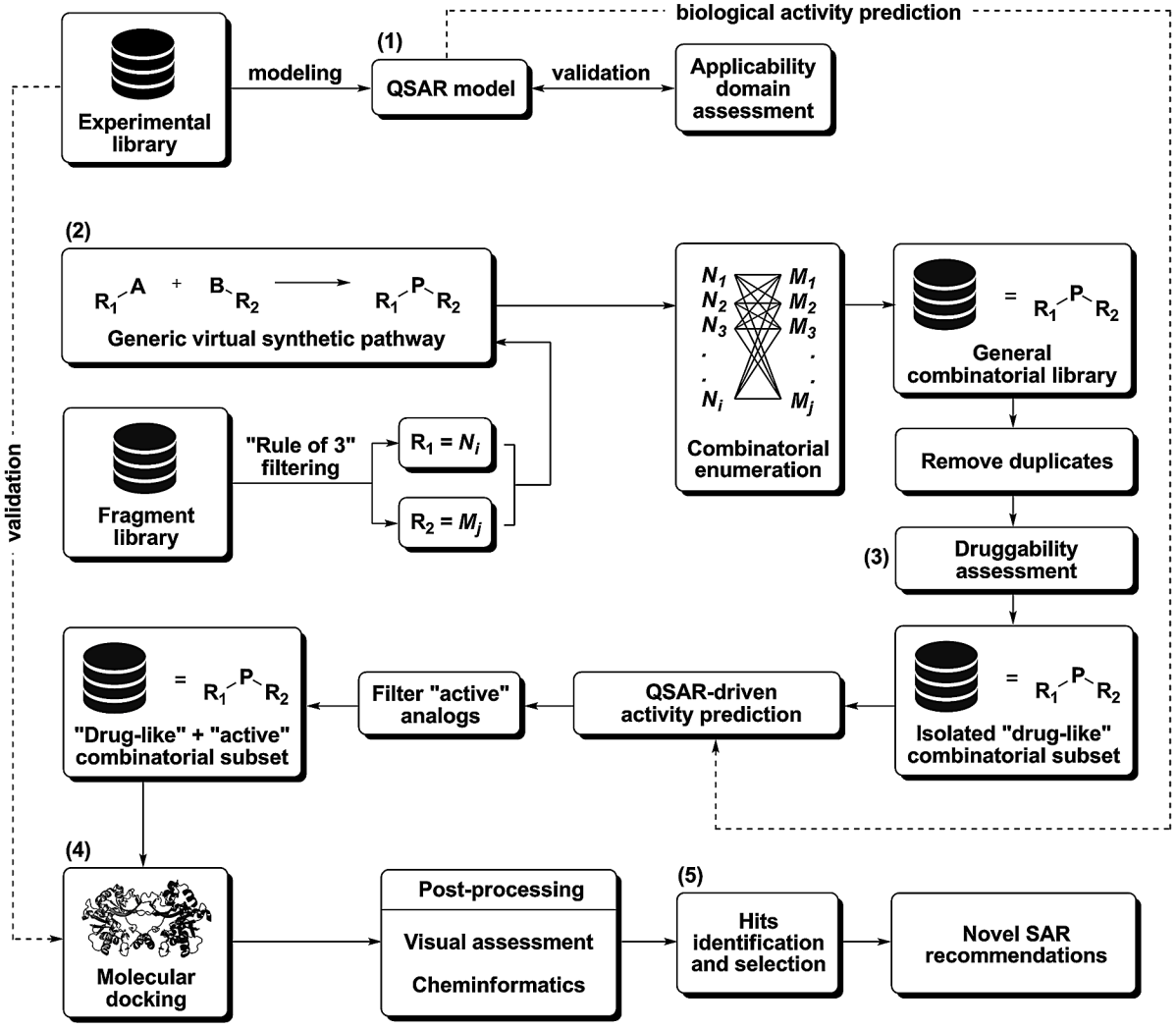
etc.). Numerous examples demonstrating a high successful rate when using such combined approaches are available elsewhere in the literature and some of the recent are summarized in Table 1.

The current state-of-the-art in the *in silico* design of novel drug candidates includes more powerful, large-scale, integrative methods, which take into consideration various drug design aspects specifically constructed to operate in a simultaneous fashion. The workflow and automation technologies available today are capable for concurrent, complex processing of billions of compounds in an astonishingly fast manner (KNIME[1], Accelrys Pipeline Pilot[2]; Warr, 2012). Based on an elegant and simple idea to visually connect a plenitude of available pre-constructed open-source and commercial component collections (e.g., CDK[3], RDKit[4], Indigo[5], Enalos[6], HCS-Tools[7], SeqAn[8], Molecular Networks[9], BioSolveIT[10], Cresset[11]) into powerful, effective, and fully customizable data pipelines, these scientific data workflow systems were initially developed for solving various chem- and/or bioinformatic problems (e.g., Taverna[12]; Kuhn et al., 2010; Truszkowski et al., 2011). However, over the years much broader scientific fields are covered and this trend is growing constantly. Nowadays, the breakthroughs in the high-performance computing also enable virtualization of the traditional HTS methods through implementation of high-throughput virtual docking (HTD) experiments on multiple various biological targets (Ellingson et al., 2013; Ellingson, Dakshanamurthy et al., 2013). Moreover, the recent explosion in the internet-based cloud technologies allow not only cloud-deposition, integration, and global accessibility of these cutting-edge *in silico* drug design solutions (Ellingson & Baudry, 2012; Hsu et al., 2013), but also facilitate the complete drug discovery process *via* simplified, shareable, collaborative research (InhibOx[13], ScienceCloud[14]).

This chapter illustrates a useful, small-scale integrated *in silico* screening approach for fast and efficient design of novel drug candidates. For this purpose, various mindfully selected 2D/3D *in silico* methodologies are covered and their proficient integration into a powerful knowledge-based virtual screening protocol is discussed. Special emphasis is given to the data flow and their gradual knowledge-based evolvement from a simple 2D to 3D environment followed by a rational concomitant data reduction for final selection of hit candidates.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 2. Graphical overview of our proposed integrated in silico screening platform for design and identification of novel hit candidates



QSAR-AIDED *IN SILICO* INTEGRATIONS FOR SMALL-SCALE LIGAND SCREENING

As reviewed so far, the integrated *in silico* approaches for small-scale ligand screening are probably the most frequently exploited. To demonstrate their practical utilization for selection of promising hit candidates from a large number of possibilities (e.g., a combinatorially-generated compound library), we propose here an effective virtual screening platform (Minovski et al., 2012; Minovski et al., 2013) by integrating various *in silico* drug design methodologies (Figure 2).

As represented in Figure 2, one can perceive that the overall *in silico* screening strategy is fundamentally constructed of five crucial interconnected levels:

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

1. Predictive QSAR modeling.
2. Virtual combinatorial library design by SAR-based fragments examination.
3. Property-based druggability assessment and QSAR-driven construction of a focused "drug-like" combinatorial library.
4. Structure-based virtual screening and post-processing of QSAR prioritized "drug-like" combinatorial compounds.
5. Hits identification, selection, and novel SAR recommendations.

The main idea behind such an integrated screening concept lies in the generation of a small molecular subspace, which can be regarded as a part of the available chemical space as well as its subsequent knowledge-based gradual reduction for identification of novel hit molecules. Nevertheless, it was estimated that the entire chemical space, which can be regarded as an ensemble of all possible molecular entities contain around 10^{60} biologically-relevant molecules that form the so-called "drug-like" chemical space (Raymond et al., 2010; Deng et al., 2013). Consequently, the objective is not to explore such a large molecular pool, but rather to generate, identify, isolate, and select a small, discrete molecular subspace that fulfill our needs – interaction with the biological system under study (Hopkins, 2008). Therefore, a crucial question arises: Are we capable to create such a small "drug-like" molecular subspace that will serve as a source for hit(s) mining? The answer to this question resides in the distinction between two substantial aspects:

1. Generation of an isolated combinatorial chemical space comprised of all possible structural analogs (e.g., novel compounds under investigation) through implementation of the virtual combinatorial library design approach, and
2. Encirclement and extraction of its embedded "drug-like" combinatorial subspace by sequential *in silico* knowledge-based evaluation supported by property-based filtering, QSAR-driven activity prediction, and structure-based virtual screening methodologies.

If carefully done, these two aspects could not only enable the generation of a high-quality "drug-like" chemical library, but could also significantly contribute to a favorable outcome at the end of the screening strategy.

Predictive QSAR Modeling

One of the earliest, but still popular and useful computational strategies in the modern drug discovery is assuredly the methodology of establishing a quantitative relationship between the chemical structures for a given compound's class and their experimentally-determined property values (Katritzky et al., 1997), i.e., an *in silico* modeling (statistical) technique commonly known as a quantitative structure-property relationship (QSPR). In the scope of the medicinal chemistry, the term "property" usually refers to an experimentally measured biological activity (e.g., IC_{50}) or even toxicity (e.g., TD_{50}) value and consequently the acronym QSPR could be referred as QSAR (quantitative structure-activity relationship) or QSTR (quantitative structure-toxicity relationship). It was found that the concept of modeling the structure-activity relationship (SAR) of chemicals in a quantitative manner is of vital importance in various drug discovery segments (Gussio et al., 1996; Ekins et al., 2002; Lee et al., 2004; Kuz'min et al., 2007). Regardless of its wide application for plain numerical predictions of biological activities or

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

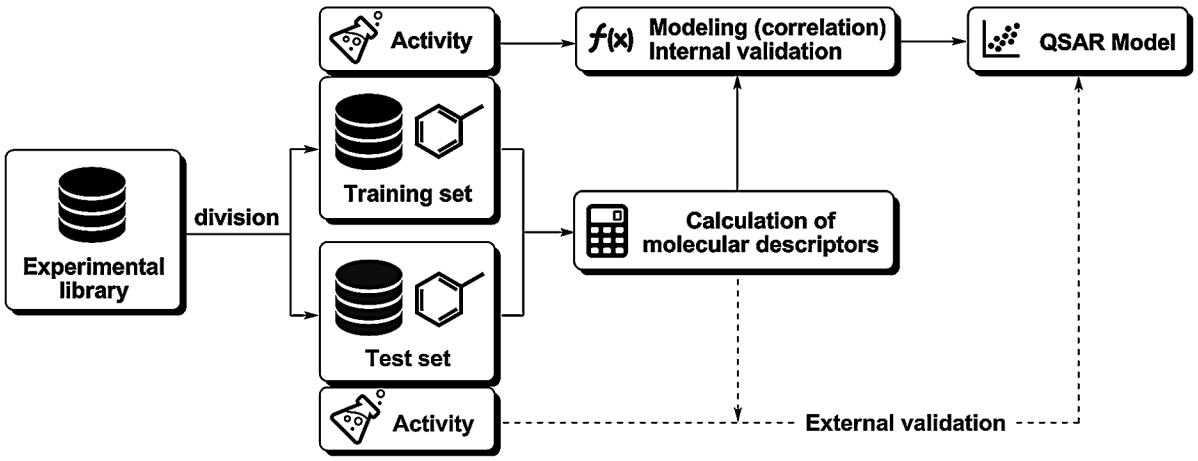
physicochemical properties for novel compounds, mechanistic interpretation of their chemical and/or biological nature encoded in a form of molecular descriptors, or even understanding the physicochemical features influencing the compound's biological response, nowadays QSAR could be recognized as a viable approach within the framework of the modern multidimensional lead optimization (Lewis, 2005; Gedeck & Lewis, 2008).

Conceptually, QSAR is based on the essential principles of medicinal chemistry, by which the biological activity of a compound is in relation to its molecular structure. As a consequence, one could expect that structurally similar compounds may have similar biological activities (Esposito et al., 2004). In QSAR, the structural features of the compounds under study are usually encoded in a numerical form commonly known as molecular descriptors. Therefore, QSAR could also be regarded as a mathematical relationship between the calculated molecular descriptors (independent variables) and *in vitro* measured biological activity values (dependent variables) for a set of known molecules. The obtained result is commonly expressed in a form of predictive mathematical model, which could readily be utilized for estimation of the biological activity values for novel, not yet synthesized compounds (Figure 2).

When QSAR as a paradigm was introduced for the first time in its entirety in the middle of the previous century (Free & Wilson, 1964; Hansch & Fujita, 1964), a plethora of QSAR approaches have been devised. Except the traditional 2D-QSAR methods (e.g., the popular Free-Wilson and Hansch-Fujita models), which are still very useful, a further breakthrough in the QSAR progress was the introduction of 3D-QSAR methods (Cramer et al., 1988) such as the comparative molecular field analysis (CoMFA) and its relative, the comparative molecular similarity indices analysis (CoMSIA). Since then, the development of QSAR approaches drastically evolved and several multidimensional QSAR congeners (e.g., 4D-, 5D, and 6D-QSAR approaches) were introduced recently (Lill, 2007; Vedani et al., 2005).

No matter which modeling method is employed, the general procedure for construction of a statistically-reliable predictive QSAR model is assembled of several consecutive steps: dataset preparation, dataset division on the so-called training and test set, calculation of molecular descriptors for both sets, modeling on the training set compounds with a simultaneous internal validation and selection of statistically-significant molecular descriptors, and finally assessment of the predictive power of the constructed QSAR model for estimation of the biological activity values for previously excluded test set compounds (Figure 3).

Figure 3. A general QSAR methodology flowchart



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

As shown in Figure 3, only training set compounds are considered in the construction of the QSAR model. In a mathematical sense, the aim is to establish a correlation between the chemical structures of the investigated compounds (training set objects) expressed in a form of calculated molecular descriptors and their corresponding biological activity values (Equation 1).

$$pA = c_0 + c_1 d_1 + c_2 d_2 + c_3 d_3 + ... + c_i d_i \tag{1}$$

where *pA* denotes the biological activity *A* in the series expressed as a negative decade logarithm, c_i are coefficients (the fitted parameters), while d_i are the calculated molecular descriptors for each compound *i* comprising the training set. For the purpose of modeling, different correlation methods (modeling algorithms) are available (Golbraikh & Tropsha, 2002), which can be generally divided in two major categories: linear modeling methods (e.g., multiple linear regression (MLR), partial least squares (PLS), etc.) and non-linear modeling methods (e.g., *k*-Nearest Neighbors (*k*NN), artificial neural networks (ANN), support vector machines (SVM), etc.). Moreover, the squared correlation coefficient is frequently used as a measure for the quality of the established predictive QSAR model, which can be calculated as follows (Equation 2):

$$R^2 = 1 - \frac{\left[\sum_{i=1}^{n_{tr}} \left(\hat{y}_i - y_i\right)^2\right] \Big/ n_{tr}}{\left[\sum_{i=1}^{n_{tr}} \left(y_i - \overline{y}_{tr}\right)^2\right] \Big/ n_{tr}} \tag{2}$$

where n_{tr} indicates the total number of objects (compounds) in the training set, $\hat{y}_i$ and y_i are the predicted and experimentally-determined (observed) biological activities for the *i*-th compound in the training set, respectively, while $\overline{y}$ designates the average of the observed values.

Another important aspect, which directly determines the predictive quality of the established QSAR model, refers to its properly performed validation (internal and external). It has been demonstrated that during the modeling, if the number of independent variables (molecular descriptors) is comparable to or higher than the number of objects (training set compounds), the probability to encounter a chance correlation between the experimental and predicted biological activity for the series significantly increases (Topliss & Edwards, 1979). In order to avoid such undesirable events, a validation of the established QSAR model is required and nowadays various validation criteria are proposed (Consonni et al., 2009; Consonni et al., 2010; Chirico & Gramatica, 2011; Roy et al., 2013). During the modeling, some commonly accepted internal validation techniques such as cross-validation leave-one-out (CV LOO) or cross-validation leave-many-out (CV LMO) are frequently utilized for assessing the statistical stability of the QSAR model (Wold, 1991). The estimation parameter carrying the statistical stability of the QSAR model is expressed as a cross-validated squared correlation coefficient (usually designated as R^2_{cv} or Q^2), which can be calculated by implementation of the same formula above (Equation 2).

In addition to the internal model validation, the robustness of the established predictive QSAR model should also be assessed. In that context, a broadly accepted methodology known as *Y*-randomization is commonly applied (Wold & Eriksson, 1995). The method is grounded on re-building of the QSAR

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

model several times (usually around ten or even more) by iterative shuffling (randomizing) the values comprising the vector of dependent variables *Y* (e.g., the biological activity values) within the original data matrix, while the independent variables (e.g., molecular descriptors used in the construction of the original QSAR model) remain intact. Therefore, one should expect that the newly established randomized QSAR models have significantly lower statistical parameters (R^2 and Q^2) comparing to those of the original non-randomized one – a result that clearly demonstrates the significance and robustness of the constructed predictive QSAR model. In case this requirement is not satisfied, it generally indicates that the established QSAR model is obtained by chance and consequently could be considered as a model of a questionable quality for its further application.

The predictive QSAR model thus established and properly validated is now ready for testing of its external predictive performances on previously excluded test set compounds, which were not used during the model development (external validation of the model). For this purpose, the external validation quantifier ($Q^2_{(te/ext)}$) is determined that reflects the external predictability of the established QSAR model and can be calculated by using the following equation (Equation 3):

$$Q^2_{(te/ext)} = 1 - \frac{\left[\sum_{i=1}^{n_{(te/ext)}} \left(\hat{y}_i - y_i\right)^2\right] \Big/ n_{(te/ext)}}{\left[\sum_{i=1}^{n_{tr}} \left(y_i - \overline{y}_{tr}\right)^2\right] \Big/ n_{tr}} \tag{3}$$

where n_{tr} is the total number of training set objects, while $n_{(te/ext)}$ designates the total number of test, i.e., external validation set objects, respectively.

Alongside the validation, the assessment of applicability domain (AD) for the established QSAR model is another important aspect, which must be taken under consideration as well (Figure 2). It directly reflects the model's reliability for its further practical application (Eriksson et al., 2003). As stated by the OECD principles for validation of the QSAR models for regulatory purposes, the QSAR model should be used within the boundaries of its clearly defined AD (OECD, 2004). To date, several approaches for estimation of the AD for QSAR models have been introduced (Zhang et al., 2006; Sahigara et al., 2012; Minovski, Župerl et al., 2013), which implementation generally vary on the modeling routine employed (linear or non-linear). Once the predictive QSAR model is rigorously validated and its robustness, predictability, and reliability are properly established, it could be readily extrapolated for estimation of the biological activity values for novel compounds (Figure 2).

Virtual Combinatorial Library Design by SAR-Based Fragments Examination

The methodology of compound library design usually refers to the generation of a list of molecules that could be synthesized utilizing either solid- or solution-phase combinatorial chemistry approach (Chabala, 1995; Coe & Storer, 1999). Traditionally, the synthetic combinatorial chemistry alias high-throughput synthesis was the major production "device" for boosting the chemical space not only in a quantitative manner, but also in the terms of its structural diversity. As mentioned previously, this concept was generally based on the hypothesis: the bigger and more diverse is the generated chemical space, the higher is the probability to identify a potential lead by high-throughput screening (HTS). Unfortunately, such

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

a strategy was disappointing for the pharmaceutical industry, since very frequently resulted in an unsatisfactory cost-benefit ratio (Lahana, 1999; Ramesha, 2000; O'Driscoll, 2004). Over the years, the drug discovery practice has evolved in a direction of swapping such unreasonable strategies with novel, cost-effective, and smart knowledge-based approaches, and the methodology of virtual combinatorial library design (known as combinatorial enumeration) was found as a suitable one (Aronov, 2002). Nowadays, by using this *in silico* approach, one can rapidly and efficiently generate thousands of novel compounds in a virtual environment that could effortlessly be manipulated.

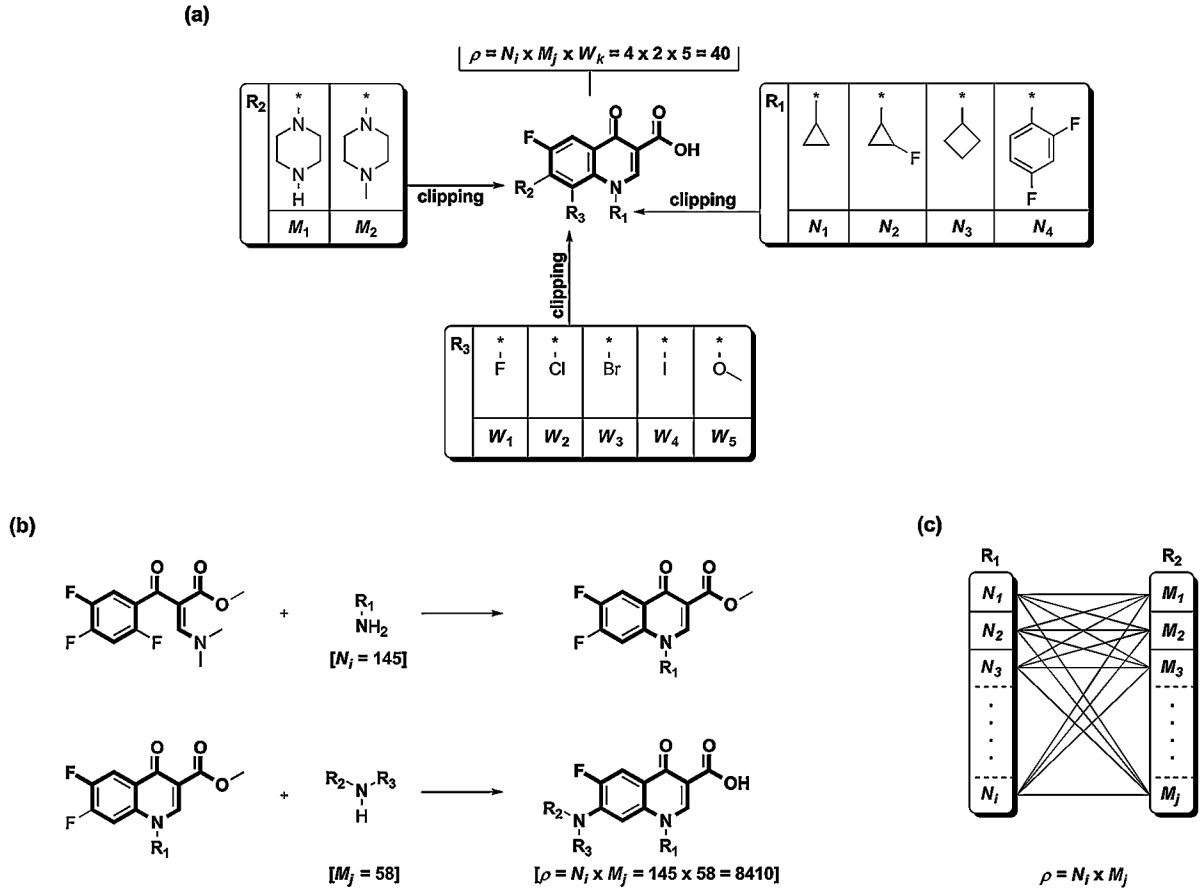
Two generally accepted strategies are commonly used for virtual enumeration of combinatorial libraries such as the Markush enumeration approach (Leland et al., 1997) as well as the reaction-based enumeration approach (Leach et al., 1999; Lobanov & Agrafiotis, 2002). In the Markush enumeration, the combinatorial library is characterized by using a Markush structural representation (widely known as "*Core plus R-groups*" molecular representation) that is the main molecular scaffold (the structural core of the combinatorial library) with one or more defined variable attachment points expressed as R-groups (e.g., R_1-, R_2-, R_3-, etc.) and a set of reagents that should be "*clipped*" (transformed) into an appropriate set of substituents (Figure 4a). On the other hand, in the reaction-based enumeration, the combinatorial library is defined by using one- or multi-step generic reaction pathway where the variable entities are represented as reactants (e.g., R_1-OH, R_2-NH_2, etc.) as illustrated in Figure 4b. Which one of these two approaches will be used, mainly depends on the complexity of the problem as well as the user's preferences (Agrafiotis et al., 2002). For instance, if the synthetic pathway is long and complex, and the scaffold representing the final product is clearly defined (no significant inter-reaction scaffold changes exist), then the Markush enumeration approach would be appropriate. Conversely, if the scaffold is variable and must be progressively constructed during the virtual synthesis, then the reaction-based enumeration approach may be the method of choice.

A key part of both enumeration approaches that could directly influence the quality as well as diversity of the generated combinatorial library is the selection of synthetically feasible reactants (e.g., fragments, building-blocks) to be attached at the pre-defined variable scaffold positions. For this purpose, a variety of *in-house* developed or commercial reactants databases are available (e.g., Accelrys ACD[15]). Nevertheless, for the sake of medicinal chemistry, a viable approach is to pre-screen the reactants database by using cheminformatically-based fragment-likeness filters such as the widely known "*Rule of 3*" ("Ro3") filtering set (MW $\leq$ 300 and *n*HBD, *n*HBA, *n*RB $\leq$ 3), where MW is the molecular weight, while *n*HBD, *n*HBA, and *n*RB, designate the number of hydrogen bond donors, hydrogen bond acceptors, and rotatable bonds, respectively (Congreve et al., 2003). This procedure assures the location of the obtained fragments within the "lead-like" chemical space (Hann & Oprea, 2004). Thus pre-treated, the obtained fragments are routinely assembled into fragment libraries, which are now ready to use in the next step – fragments selection for combinatorial enumeration.

The fragments selection is frequently done by substructure search (SSS) procedure enforced by the fragment(s) designation. For instance, the R_1-OH designation determines the SSS-supported retrieval of all organic fragments from a fragment library that contain –OH group(s) including alcohols, phenols, carboxylic acids, etc. However, if one takes into account all the fragments retrieved by SSS without their subsequent examination, then the generated combinatorial library would be so diverse so it can be compared to the chemical space obtained by the traditional high-throughput synthesis that is not fully in agreement with our expectations. In order to avoid this, the retrieved fragment subsets should be carefully examined (automatically and visually, if possible). In practice, the fragments examination is usually SAR-based, and therefore the availability of any SAR knowledge for the molecular entity (e.g.,

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 4. Combinatorial enumeration approaches (examples of 6-fluoroquinolones combinatorial libraries): (a) Markush enumeration, (b) reaction-based enumeration, and (c) schematic representation of a combinatorial enumeration process for a simple reaction system configured of two subsets of reactants R_1 (N_1…N_i) and R_2 (M_1…M_j)



a drug with well-established potency) used as a template for analogs design would be of high significance. Moreover, the SSS procedure can also be additionally employed for efficient elimination of those fragments with undesirable reactive functionalities and thereby the risk for late determination of *in vitro* false positives could be minimized (Rishton, 1997; Rishton, 2003). The fragment subsets prepared by using this way can finally enter the combinatorial enumeration process.

The combinatorial enumeration (virtual combinatorial generation) can be regarded as a straightforward procedure of statistical non-repetitive fragmental permutation where each fragment of a given reactants subset "interact" with each fragment of the other reactants subsets within the previously defined molecular scaffold (Figure 4c). Let's consider a simple reaction system configured of two subsets of reactants R_1 (N_1…N_i) and R_2 (M_1…M_j). Following the combinatorial enumeration process (Figure 2 and Figure 4c), one could easily estimate the total number of combinatorial analogs obtained by the enumeration process as a multiplication product between each pair of reactants (N_1…N_i and M_1…M_j). Mathematically, this process can be expressed by using the following equation (Equation 4):

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

$$\rho = N_i \times M_j \tag{4}$$

where N_i are the total number of fragments within the R_1 subset, M_j are the total number of fragments within the R_2 subset, while ρ is the total number of products (combinatorial analogs) obtained by the combinatorial enumeration (Wieland, 1997). Furthermore, an additional check of the constructed virtual combinatorial library for duplicate molecular entities and their subsequent removal is frequently a good practice (Figure 2).

Property-Based Druggability Assessment and QSAR-Driven Construction of a Focused "Drug-Like" Combinatorial Library

As exemplified above, the methodology of virtual combinatorial library design resulted in obtaining a general virtual combinatorial library comprised of all possible combinations of structural analogs for the compound(s) under study (Figure 2); therefore, it can be regarded as an isolated part of the available chemical space (Bohacek et al., 1996). However, the essential question here is, whether this isolated molecular pool is useful (at all) for further selection of novel drug molecules?

Previously, we demonstrated that the so-called "drug-like" chemical space was estimated to contain approximately 10^{60} biologically relevant molecules with MW ~ 500 (Raymond et al., 2010; Deng et al., 2013). Thus, one should primarily assess whether our isolated virtual combinatorial library represents a part of it. Moreover, it was found that the combinatorial algorithm in first instance increases the molecular complexity (Hann et al., 2001); as a consequence, one should also bear in mind the low probability that each generated compound within such a virtual combinatorial library possess "drug-like" properties (Walters et al., 1999; Muegge, 2003). To eliminate these dilemmas as well as to identify and distill those compounds representing the "drug-like" chemical space, the constructed general virtual combinatorial library should be further subjected to a thorough property-based druggability assessment (Barril, 2012).

It was found that the "drug-like" properties confer favorable ADMET (absorption, distribution, metabolism, excretion, and toxicity) attributes to a compound. However, it is important to clarify which are those properties that describe a single molecule as a potential drug? In a broader sense, the "drug-like" properties are defined as intrinsic molecular features that can strongly influence the optimization of their pharmacological characteristics (Borchardt, 2004). During the last two decades, various attempts were made for identification of these molecular properties and clarification of their role. Owing to the seminal work of Christopher A. Lipinski who pioneered the famous "Lipinski's rule of 5" ("Ro5"), the concept of "drug-likeness" acquired a totally new dimension in the drug discovery world (Lipinski et al., 1997). It states that the probability for a compound to be poorly absorbed after oral administration increases if any two of the following rules are violated:

- Molecular weight (MW) is less than 500 Da.
- Number of hydrogen bond donors (*n*HBD: OH/NH groups) is equal or less than 5.
- Number of hydrogen bond acceptors (*n*HBA: O/N) is less than 10.
- Calculated logP is less than 5.0 (by using ClogP) or 4.15 (by using MlogP).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Nevertheless, an important aspect to be taken with precaution when using the "Ro5" as "drug-like" filter for druggability assessment is the lipophilicity parameter (logP). Namely, in the "Ro5" druggability filter the logP parameter refers to the lipophilicity of a compound in its neutral state. On the other hand, the majority of the drug molecules (~ 95%) prevail in an ionizable state. Since the orally administered drugs are mainly absorbed in the small intestine where the environment is slightly acidic, the pH parameter should also be considered. Therefore, a modified "Ro5" concept was introduced recently, favoring the implementation of the pH dependent version of logP at intestinal pH ~ 5.5 ($logD_{5.5}$), instead of the classical logP parameter for estimation of the compounds lipophilicity (Bhal et al., 2007).

Another critical molecular parameter that strongly correlates with the compounds membrane permeability and consequently their oral bioavailability is the polar surface area (PSA) designated as a sum of the van der Waals surface areas of the polar atoms (O and N) within the molecule. Accordingly, Daniel F. Veber assembled a useful two-parameter rule set, i.e., "*Veber rule*" by which a molecule is likely to have a good bioavailability after oral administration if the following criteria passed (Veber et al., 2002):

- PSA is equal or less than 140 $Å^2$ (or ≤ 12 *n*HBD + *n*HBA).
- Number of rotatable bonds (*n*RB) is equal or less than 10.

In addition to these "drug-like" filters, there are also some other similar rule sets carrying various drug-discriminating physicochemical properties including the Pardridge blood-brain barrier (BBB) permeability filter (Pardridge, 1995), Clark-Lobell's BBB permeability filter (Clark, 2003; Lobell et al., 2003), the Pharmacophore Point Filter (Muegge et al., 2001), and many others. It is important to stress out that these *in silico* filters should not be used indiscriminately, i.e., there must be a solid ground why to use one filter, and not another. For instance, if one needs to screen a set of intravenously administered compounds for their BBB penetrating abilities, then one of the available *in silico* BBB permeability filters should be applied (e.g., Clark-Lobell's BBB filter), and not the filters for prediction of intestinal absorption (e.g., Lipinski's Ro5). Once the appropriate "drug-like" filter is selected, it can be proficiently employed for *in-depth* screening of our previously constructed general virtual combinatorial library. Those combinatorially-generated compounds (if any) which successfully pass all the pre-defined filter's criteria represent our so-called isolated "drug-like" combinatorial subspace.

The selected compounds are further subjected to the previously constructed predictive QSAR model (Figure 2) to obtain *in silico* prediction of the biological activity values. After sorting the compounds by decreasing the predicted biological activities, the priority list could be further reduced if some activity range data are available (usually obtained by *in vitro* functional and/or biochemical studies of known drugs). This could be considered as an additional filter for extraction of those "drug-like" combinatorial analogs with predicted biological activity values within the desired pre-defined activity range. For example, the *in vitro* inhibition assays of 6-fluoroquinolone (6-FQ) antibacterials against *M. tuberculosis* DNA gyrase enzyme revealed that these inhibitors are active in the range of $0.0 \leq$ MIC [μg/mL] $\leq$ 1.0-2.0, where MIC refers to the minimal inhibitory concentration. Therefore, one can apply this activity range as *in silico* filter for selection of those combinatorially-generated 6-FQs which QSAR-predicted MIC values fall within the pre-defined MIC boundaries (predicted "active" 6-FQ analogs). The obtained combinatorial subset comprised of all "drug-like" plus "active" (as predicted by the QSAR model) combinatorial analogs can be regarded as a focused "drug-like" combinatorial library which is now prepared to enter the structure-based virtual screening stage of our *in silico* integrated protocol (Figure 2).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Structure-Based Virtual Screening and Post-Processing of QSAR Prioritized "Drug-Like" Combinatorial Compounds

The *in silico* strategies elaborated so far generally belong to the class of ligand-based methodologies which usefulness and versatility in various drug discovery segments are broadly reviewed (Vidal et al., 2011). Notwithstanding their mounting success, a major drawback that follows all these approaches resides in their incapability to account for the possible intermolecular interactions between a given ligand and its biomolecular target. In that context, the structure-based virtual screening approaches were found as particularly important to proficiently complement the missing gap (Ghosh et al., 2006; Kroemer, 2007; Villoutreix, et al., 2009) and here the methodology of automated molecular docking was recognized as a pivotal one in the framework of the modern *in silico* drug discovery (Lybrand, 1995; Morris & Lim-Wilby, 2008).

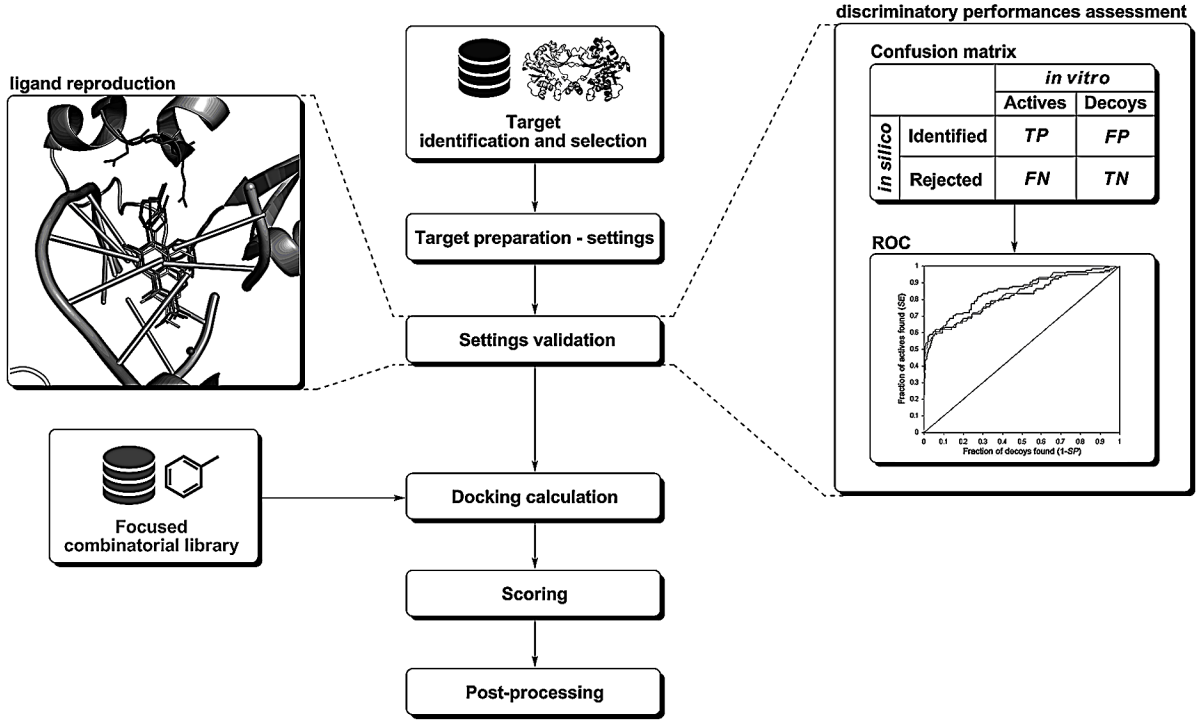
As mentioned previously, the availability of 3D structural information for the biological target at atomic resolution (preferably ≤ 2.5 Å, obtained by experimental techniques such as X-ray crystallography or NMR) is critical for its implementation as virtual ligand filter. Herein, the generic term "biological target" usually refers to a biomacromolecular structure (e.g., a protein, nucleic acid, or their complex) into which the ligand (e.g., a small drug molecule) is being docked. This concept ordinarily relates to a method commonly known as protein-ligand docking, however, during the last few years an expanding interest was observed for protein-protein docking methods as well (Gray et al., 2003; Wang et al., 2007; Moal et al., 2013). Nevertheless, for the purpose of *in silico* design of small molecular entities as possible future drug candidates, the protein-ligand docking methods were found as the most frequently exploited.

As illustrated in Figure 5, a key part of the protein-ligand docking methods is the target identification and its subsequent selection. This imposes a profound understanding of the nature of the biological system under study including its function and mechanism in physiological or pathophysiological purposes, the location of the catalytic and/or binding site(s), and preferably the ligand-binding mechanism if available (Seifert & Lang, 2007). Moreover, one should be aware of some possible scenarios that could frequently follow the target selection and later can strongly affect the final epilog of the docking study. Ideally, the availability of a protein-ligand complex structure originating from the designated organism is of invaluable importance in the protein-ligand docking studies. This often enables a successful run of the entire docking process in a straightforward way, and consequently obtaining a productive outcome at the end of the screening strategy. Unfortunately, this is not always a case. Some protein targets are not or cannot be co-crystallized with the ligand, even if they originate from the desired organism. There are also some cases where the structure of the protein-ligand complex is available, but originates from other species, or the desired protein structure is not available as a whole, but as separated subunits. In such cases, some alternative approaches in the scope of the target preparation stage could be applied including automated homology modeling and/or ligand's binding site recognition (Xiang, 2006; Minovski et al., 2013; Yang et al., 2013).

Once the designated biological target is identified and selected, it should be properly prepared for virtual screening. Except some preparation routines related to the entire target structure such as adding hydrogen atoms and elimination of the water molecules, the target preparation stage usually refers to the settings of the binding site for docking calculation. This often includes assignment of the protonation, stereoisomeric, and tautomeric states of the amino acid residues covering the binding pocket (Anderson, 2003). Additionally, small molecular entities (e.g., water molecules, metal ions, and other co-factors) present in the binding pocket should also be removed, unless there is a strong evidence for

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 5. A typical methodology workflow covering the essential steps of a protein-ligand docking



their specific role in the protein-ligand binding mechanism (Cheng et al., 2012). For instance, the water molecules present in the binding pocket could serve as hydrogen bonding bridges between the ligand and surrounding amino acid residues. Furthermore, if a co-crystallized ligand conformation is available, it can be easily employed for assignment of the screening area, i.e., the ligand search space that is usually a sphere-shaped constrain with a centroid defined by the Cartesian coordinates (x, y, z) of a selected ligand's atom. Otherwise, the screening area can algorithmically be assigned as an integrated part of the binding site recognition routine.

It should be stressed, however, that in cases where a protein-ligand complex is available (experimentally determined or assembled by homology modeling), an initial docking validation run is required. This involves reproduction of the spatial orientation and conformation for the co-crystallized ligand that directly reflects the quality of all the settings performed previously. The ligand reproduction assessment usually consists of re-docking of the co-crystallized ligand conformation and its subsequent comparison to the calculated docking solutions. As judging criteria for the performed ligand reproduction, the all-atom or heavy-atoms root-mean-square deviation (RMSD) value between each calculated docking pose and the co-crystallized ligand conformation is usually calculated. The calculated RMSD values below 2.0 Å are generally accepted and often indicate a successfully performed ligand reproduction (Verdonk et al., 2003). Moreover, if a library comprised of compounds with experimentally-determined *in vitro* biological activities is on hand (e.g., similar to the one used for QSAR modeling), then it could be proficiently employed in an additional docking validation experiment for evaluation of the target's discriminatory performances – a valuable guideline that indicates the capability of our biological target to effectively discriminate between known active and inactive ligands (Figure 2). It was demonstrated that

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

such a validation is of exceptional importance where one needs to rank two or more analogous protein homology models for their ligands discriminating capabilities (Minovski et al., 2013). In that sense, a widely accepted strategy is the so-called *enrichment* of the library of known actives with an incomparably higher number of decoy molecules (Huang et al., 2006). Here, a decoy is defined as a molecular entity that physicochemically matches the known active compounds, i.e., shares the same physicochemical features, but at the same time is topologically distinct. The main idea behind this concept is to evaluate how good a constructed protein homology model can identify the known actives in the wave of decoy molecules. For that purpose, the "receiver operating characteristic" (ROC) curve methodology proved to be the most suitable one for solving binary classification problems (Triballeau et al., 2005; Hevener et al., 2009). The ROC curve covers the sensitivity (*SE*) and specificity (*SP*) of the model, where *SE* is graphically expressed as a function of (1-*SP*). These two features one could easily determine from the confusion matrix representing a binary classification (Manallack et al., 2002). Let's consider a case where a docking-based virtual screening protocol was applied for evaluation of the discriminatory power of two analogous protein homology models to correctly identify active ligands among a large pool of decoy molecules. In the context of binary classification, four classes could be distinguished: true positives (*TP* = *in vitro* actives and *in silico* identified), true negatives (*TN* = *in vitro* inactives and *in silico* rejected), false positives (*FP* = *in vitro* inactives and *in silico* identified), and false negatives (*FN* = *in vitro* actives and *in silico* rejected). Consequently, *SE* and *SP* parameters for both homology models can be calculated as follows (Equation 5 and Equation 6):

$$SE = \frac{TP}{TP + FN} \tag{5}$$

$$SP = \frac{TN}{TN + FP} \tag{6}$$

where *SE* is defined as the ratio of *in silico* correctly identified actives (positives) over all compounds which are known to be truly active *in vitro*, while *SP* is defined as the ratio of *in silico* correctly identified inactive molecules (negatives) over all compounds which are known to be truly inactive *in vitro*. Finally, the area under the ROC curve (ROC-AUC) is computed directly from the ROC plot – a convenient metric that reflects the discriminatory performances of a protein homology model as a virtual ligand filter (Equation 7).

$$\text{ROC-AUC} = 1 - \frac{1}{N_A}\sum_{i=1}^{N_A}\frac{N^*_{i,D}}{N_D} \tag{7}$$

where N_A represents the total number of actives, N_D is the total number of decoy molecules, while $N^*_{i,D}$ depicts the number of decoy molecules that are higher ranked than the *i*-th active compound.

The ROC-AUC values can vary between 0 (poor model's discriminatory performances) and 1 (excellent model's discriminatory performances), whereas ROC-AUC = 0.5 corresponds to a random selection (Triballeau et al., 2005). Therefore, the higher values for ROC-AUC (preferably above 0.5), pinpoint to better discriminatory performances of a given protein homology model, and *vice versa*. Nevertheless, while ROC-AUC metric evaluates the overall performance of a given protein homology model considering the entire data (actives and decoys), it is not capable to effectively account for the *early enrichment* of active molecules in small portions of the entire compound library (e.g., 0.5%, 1.0%, or 2.0%). Therefore,

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

to fulfill the missing gap an additional metric commonly known as *enrichment factor* ($EF_{x\%}$) should be applied, which is derived from the ROC curve as well (Jahn et al., 2011). The $EF_{x\%}$ for a given library portion ($x\%$) can be calculated as follows (Equation 8)

$$EF_{x\%} = \frac{\frac{N_A^*}{N_{x\%}}}{\frac{N_A}{N_A + N_D}} \tag{8}$$

where N_A is the total number of active molecules in the entire library, N_D is the total number of decoy molecules in the entire library, $N_{x\%}$ is the total number of compounds in the observed portion of the library ($x\%$), and N^*_A represents the total number of actives found within the observed portion (Jahn et al., 2011). Thus validated, the target can now be used for performing molecular docking calculations on novel compounds such as those that comprise our previously constructed focused "drug-like" combinatorial library derived by QSAR-driven prioritization (Figure 2).

The molecular docking calculation refers to a virtual examination of the pre-assigned ligand search space (screening area) followed by ranking of the calculated docking solutions (i.e., the generated ligand binding poses) for subsequent determination of the correct binding modes with the target (Morris & Lim-Wilby, 2008). In the attempt to find the energetically most favorable binding solutions, the search algorithm runs several times for each ligand entering the docking calculation. The number of trials per ligand is often assigned by the user and usually corresponds to the total number of calculated poses at the end of the docking calculation (frequently 3-10). Whether the calculated ligand's binding pose is energetically favorable depends on the computed protein-ligand interaction energy (known as a score) for that ligand. For this purpose, many docking tools are equipped with various scoring functions (Halperin et al., 2002; Jain, 2006; Huang et al., 2010). The scoring function initially evaluates the spatial orientation and conformation of the calculated ligand pose by calculating a simple energy function (e.g., a force field constructed of electrostatic and attractive/repulsive van der Waals energy terms), while later more complex scoring schemes are used for estimation of the ligand binding affinity (Gohlke & Klebe, 2001).

The docking calculation is considered finished once a sufficient number of solutions (docking poses) have been generated for each compound entering the docking process. The obtained results are usually stored in a compact form (e.g., a virtual compound library) where the estimated docking solutions for each docked ligand are organized as small clusters. The docking poses within each cluster are frequently ranked by the calculated scoring function in a descending order, i.e., the highest scored docking pose is located at the top of each cluster. Naturally, one would expect that these top-scored solutions are indeed the best binding poses. Unfortunately, it was found that the hit(s) selection based solely on the highest calculated dock score is not always sufficient, mainly as a consequence of the imperfection of some scoring functions to correctly account for the ligand's binding mode (Kitchen et al., 2004; Cheng et al., 2012). Therefore, irrespectively of the calculated dock scores, a post-processing of the obtained results, i.e., a thorough post-docking analysis should be accomplished (Minovski et al., 2012; Minovski et al., 2013). Thus, a good practice is to perform an *in-pocket* visual assessment of all the calculated docking solutions, which could raise the entire post-docking analysis to a significantly higher, chemically intuitive, knowledge-based level (Doman et al., 2002). From a drug design perspective, it may involve an

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

evaluation of various useful attributes such as dock pose spatial orientation, conformation, fitness of the calculated dock pose with the co-crystallized ligand, as well as an *in-pocket* SAR-based pharmacophore assessment based on the estimation of how many common pharmacophoric features are shared between an investigated dock pose and the co-crystallized ligand (Minovski et al., 2013). The latter could be recognized as particularly interesting, since it could effectively account for the potential protein-ligand interactions. Nevertheless, while the visual inspection might be feasible in cases when a reasonable number of docking solutions should be checked (e.g., several hundreds of docking poses), it might be totally impractical when one needs to assess thousands of docking poses. In such circumstances, the automated solutions to this problem can be regarded as one of the most needed. Nowadays, owing to the advancements in cheminformatics, various useful tools for automated post-docking analysis based on mapping of the desired protein-ligand interactions are developed (Marcou & Rognan, 2007; Rastelli et al., 2009; Bouvier et al., 2010). However, within the framework of an integrated protocol for small-scale ligand screening (such the one illustrated in this chapter), where the starting compound library is significantly reduced throughout all the levels (Figure 2), only a reasonably small number of compounds are subjected to docking calculation. Therefore, taking into account all the attributes mentioned above, the post-docking analysis based on visual examination of the calculated docking solutions could be equally good for final identification and selection of potential novel hits.

HITS IDENTIFICATION, SELECTION, AND NOVEL SAR RECOMMENDATIONS

As mentioned above, the post-docking analysis based on visual examination provides some useful information about calculated dock poses as well as their putative binding mechanism with the key amino acid residues wrapping the target's binding site. Nonetheless, the question that arises here is how this information might be helpful for identification and selection of the most promising hit candidates? For the purpose of hits identification, various *in silico* methods and tools are devised (Marcou & Rognan, 2007; Rastelli et al., 2009; Bouvier et al., 2010; Moldover et al., 2012; Liu et al., 2013), however, in the context of our proposed integrated *in silico* screening protocol, we introduce here a novel and simple semi-automated Boolean-based [T/F (true/false)] clustering method for hit(s) identification, covering all the attributes determined in the scope of the visually derived post-docking analysis (Minovski et al., 2013).

The Boolean-based clustering method is substantially organized in three consecutive levels:

Level 1, Geometric Properties Assessment: Spatial examination of each calculated dock pose relative to the experimental co-crystallized ligand conformation and building a cluster of (T)-signed dock poses. Within this level, a total of three geometric properties should be visually assessed: (1) spatial orientation (how a calculated dock pose is spatially oriented relative to the position of the co-crystallized ligand), (2) pose fitness (how well a calculated dock pose fits the co-crystallized ligand), and (3) number of matching pharmacophoric features (how many common pharmacophoric features are shared between a calculated dock pose and the co-crystallized ligand). The latter implies to an *in-pocket* generation of a structure-based pharmacophore model for both the co-crystallized ligand conformation and the investigated dock pose. For this purpose, various powerful automatic pharmacophore generation tools are available (e.g., Catalyst[16], LigandScout[17]). Contrary to the properties (1) and (2), which are evaluated qualitatively by using the Boolean-based (T/F) signing scheme, the third property (number of matching pharmacophoric features) is evaluated quantita-

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

tively. Whether an investigated dock pose will pass to the next level or will be rejected, depends on the positive outcome for all three geometric properties. For instance, if the calculated dock pose is properly oriented (T), fits well (T), and shares the desired number of pharmacophoric features with the co-crystallized ligand conformation (e.g., 5), then as a consensus, the dock pose could be estimated as geometrically good (T) and consequently goes further. Otherwise, the dock pose will be rejected (F). This assessment should be repeated for all the calculated dock solutions. At the end, only those poses signed as (T) should be distilled as a separate cluster and used in the second level of the Boolean-based clustering method.

Level 2, Score-Based Clustering: (T)-signing of the *Level 1* dock poses with calculated dock score within a pre-defined, desired range (e.g., highly scored dock solutions) and building a new cluster of highly scored (T)-signed hits. Those calculated dock solutions, which dock score is outside the boundaries of the pre-defined range, would be rejected (F).

Level 3, Activity-Based Clustering: (T)-signing of the *Level 2* hits with predicted biological activity values within a pre-defined, desired range (e.g., highly "active" combinatorial analogs as estimated by the previously used predictive QSAR model) and building a new cluster of (T)-signed most "active" hits. It should be stressed, however, that all the combinatorial analogs comprising our previously constructed focused "drug-like" combinatorial library are somehow hypothetically "active" as defined by the additionally implemented biological activity range filter. For instance, as previously stated, it was found that 6-fluoroquinolone antibacterials are active as *M. tuberculosis* DNA gyrase inhibitors in the range of $0.0 \leq \text{MIC}\ [\mu g/mL] \leq 1.0$ or 2.0. Let's consider that one needs to extract those highly "active" 6-fluoroquinolones with predicted $\text{MIC} \leq 0.05\ \mu g/mL$. Therefore, the dock solutions which QSAR predicted biological activity is less or equal then 0.05 μg/mL will be signed as (T) and used in the end phase of the hit(s) identification/selection process.

Finally, all (T)-signed dock poses isolated within the last level of the Boolean-based clustering method, are further subjected to an additional substructural assessment of the combinatorially attached fragments. This selection procedure resides in the determination of the most frequently appearing fragments at the pre-defined variable scaffold positions followed by a subsequent comparative SAR analysis for exposing novel possible guidelines utilizing a well-established SAR for known drugs. The combinatorial hits thus selected are now ready for synthesis, purification, and further *in vitro* biological activity verification.

CASE STUDY: *IN SILICO* DESIGN AND IDENTIFICATION OF NOVEL 6-FLUOROQUINOLONES AS POTENTIAL INHIBITORS AGAINST *MYCOBACTERIUM TUBERCULOSIS* DNA GYRASE

The integrated *in silico* screening protocol elaborated in the previous section (Figure 2) could generally be applied for the design and identification of novel drug analogs, irrespectively of the drugs class that are derived from as well as the *in silico* methodologies employed. However, to demonstrate its practical implementation, we opted to illustrate a specific example related to the design and identification of novel 6-fluoroquinolone (6-FQ) antibacterials as potential inhibitors against *M. tuberculosis* DNA gyrase enzyme.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

M. tuberculosis DNA Gyrase as a Therapeutic Target and 6-Fluoroquinolone Antibacterials

It is widely known that in all living organisms, the correct spatial DNA topology is of crucial importance for the proper regulation of the DNA processing including all the basic biological processes, control of the gene replication and transcription, as well as the DNA segregation (Wasserman & Cozzarelli 1986; Liu et al., 2009). In order to be spatially correct, the DNA topology is fundamentally maintained by a specific family of enzymes broadly known as the DNA topoisomerases (Sissi & Palumbo, 2010); this family of superior molecular nanomachines is substantially involved in the maintenance of two vital sub-cellular processes such as the DNA single-strand breaks (controlled by the so-called type I enzymes) and DNA double-strand breaks (controlled by the so-called type II enzymes).

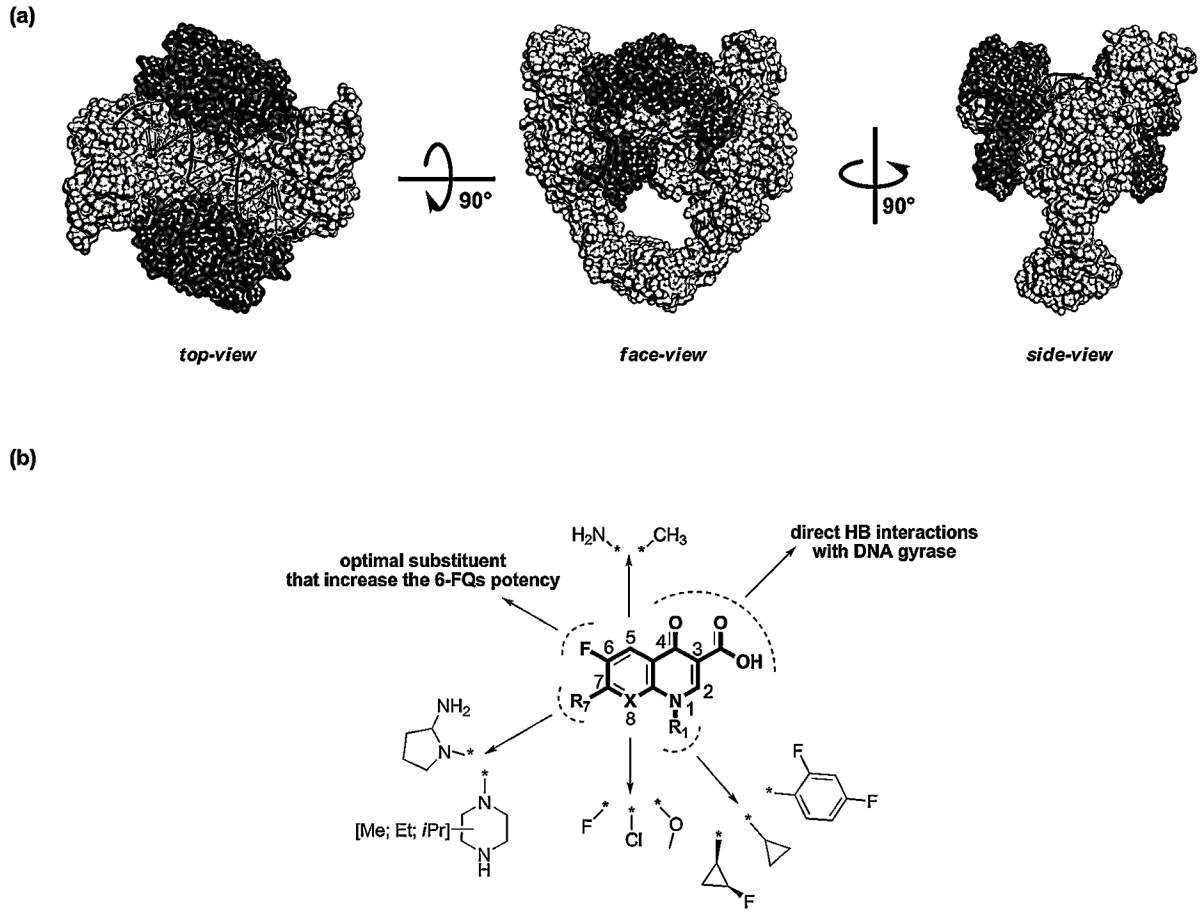
In contrast to the other, mainly higher organisms, which possess multiple topoisomerase enzymes involved in various essential sub-cellular functions, the bacterial organisms usually possess two general types of topoisomerases (Levine et al., 1998) - DNA gyrase enzyme (responsible for the unwinding of the bacterial DNA during the DNA replication phase) and topoisomerase IV (a DNA gyrase paralogous form involved in the DNA decatenation process). However, unlike other bacterial species, *M. tuberculosis* is an unusual and unique bacterial organism, since it possesses only one type II topoisomerase - the DNA gyrase with a specific simultaneous functional role of topoisomerase IV (Cole et al., 1998; Collin et al., 2011; Bouige et al., 2013). Structurally, the mycobacterial DNA gyrase (Figure 6a) is comprised of two cardinal subunits GyrA and GyrB (relevant to the ParC and ParE subunits in topoisomerase IV) that together assemble a functional heart-shaped heterotetrameric complex A_2B_2 (C_2E_2 in topoisomerase IV). The structural and biochemical studies performed so far revealed that the GyrA subunit is responsible for the mycobacterial DNA breakage/reunion catalytic process (i.e., DNA replication and elongation), while the correct spatial topology of the mycobacterial DNA is maintained by the GyrB subunit (Levine et al., 1998; Cole et al., 1998). This catalytic process in *Mycobacteria* was recognized as a promising targeting mechanism for efficient tuberculosis chemotherapy (Ferrero et al., 1994; Maxwell, 1997; Collin et al., 2011).

The quinolone antibacterials (e.g., Nalidix acid and its structurally derived analogs) were found as the sole and most efficient DNA gyrase inhibitors so far. Among them, the 6-FQ class (quinolone's structural congeners that contain a fluorine atom attached at the 6 position of the main quinolone core) are probably one of the most utilized antitubercular agents in the clinical practice (Figure 6b). Their unique mode of action is grounded on the inhibition of the essential DNA gyrase mechanisms (e.g., supercoiling and relaxation of the double-stranded DNA), followed by instant formation of an irreversible covalent complex between the GyrA subunit, 5'-end of the mycobacterial DNA, and the 6-FQ inhibitor itself (Gellert et al., 1977; Sugino et al., 1977). It was found that the complex thus formed is highly stable and cytotoxic for the *Mycobacteria*, leading to an irrecoverable distortion of the entire DNA topology, failures in the DNA processing mechanisms, and finally bacterial cell destruction (Drlica & Malik, 2003).

The current SAR knowledge of 6-FQ antibacterials (Tillotson, 1996) explicitly shows the crucial 6-FQs scaffold positions, which structural alteration could significantly enhance or reduce the antimycobacterial activity (Figure 6b). As demonstrated, the main quinolone moiety (1,4-dihydro-4-oxo-3-pyridinecarboxylic acid) is of significant importance for the antimycobacterial activity. At the position 1, the substitution of a cyclopropyl group was found as the optimal one, however, various bulky substituents could also be attached at this position, which could enhance not only the lipophilic profile of the entire drug, but could also increase its metabolic stability after possible oral administration. On the other hand, the positions 5

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 6. (a) Three-dimensional structural views of the mycobacterial DNA gyrase enzyme; the individual DNA gyrase parts are colored differently – GyrA subunit (in light gray), GyrB subunit (in dark gray), and the mycobacterial DNA molecule (in black); (b) The current SAR knowledge of the 6- FQ antibacterials



and 8 could successfully be substituted with some small substituents (e.g., -NH_2 or CH_3 at position 5, i.e., -F, -Cl, or -OCH_3 at position 8); these structural alterations could lead to an increased biological activity. The position 6 of the main scaffold is commonly reserved for a fluorine atom, which was found as the optimal one leading to an increased potency, however, it could also be replaced by some small substituents. Three positions (3, 4, and 7) on the main quinolone core were found as of exceptional importance for the 6-FQs antimycobacterial activity. The mechanistic studies revealed that the substituents attached at these positions could establish direct interaction not only with the enzyme (GyrA subunit), but also with the DNA molecule (Laponogov et al., 2009; Laponogov et al., 2010). Thus, the positions 3 and 4 should be occupied by a carboxyl and carbonyl group, respectively, as they lead to a significant enhancement of the 6-FQs potency through establishing a direct hydrogen-bonding interaction with the surrounding amino acid residues of the enzyme (Tillotson, 1996). On the other hand, the position 7 was recognized as a key attachment point for the biological activity of 6-FQ antibacterials, and aminopyrrolidinyl- or piperazinyl-like substituents were found as the most important, leading to an additional stabilization of the 6-FQ in the complex formed between the DNA gyrase enzyme and the mycobacterial DNA.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

However, as a consequence of the considerable structural divergences between the available gyrase/topoisomerase crystal structures (Laponogov et al., 2009; Laponogov et al., 2010; Wohlkonig et al., 2010; Bax et al., 2010), the explicit protein-ligand interactions mainly remain unclear, and regrettably the 6-FQs binding mechanism is still a subject of speculation. Moreover, some recently confirmed quinolone-caused amino acids alterations largely located at the GyrA α2 and α3 helices delineating a key part of the quinolone-binding pocket (QBP), were found as one of the major determinants of the ineffectiveness of the current 6-FQs in the chemotherapy of tuberculosis (Johnson et al., 2006; Shi et al., 2006; Matrat et al., 2006; Groll et al., 2009). All these issues, represent a major challenge towards the structural optimization of the existing 6-FQs or even development of novel, more effective 6-FQ antituberculars, and nowadays, the integrated *in silico* drug design approaches (Figure 2) have proven to be skillfully applied in achieving these goals.

The structural information used in this study including the description of datasets of 6-FQs, i.e., DNA gyrase/topoisomerase crystal structures as well as all the methodologies employed are broadly elaborated in our previous publications (Minovski & Šolmajer, 2010; Minovski et al., 2011; Minovski et al., 2012, Minovski et al., 2013), and therefore we give here just a short summary.

6-FQs Compound Libraries

Two compound libraries of structurally similar 6-FQs coming from different sources (named as *ExpLib* and *CombiLib*, respectively) were primarily utilized as a starting point for QSAR predictive modeling, and later on to perform structure-based VS experiments for identification and selection of novel 6-FQ antitubercular agents.

The *ExpLib* library comprised of 145 structurally similar 6-FQs with experimentally measured biological activity values expressed as minimal inhibitory concentrations – MIC_{exp} [μg/mL] was collected from an online structural data source (Division of AIDS Anti-HIV/OI/TB Therapeutic database[18]; Minovski et al., 2011). According to the 6-FQs activity information obtained within the framework of various functional and biochemical studies performed so far, one could easily distinguish between the active and inactive 6-FQs comprising our *ExpLib* library (Aubrey et al., 2006; Pantel et al., 2011; Pantel et al., 2012); therefore, 114 out of total 145 6-FQs could be determined as active inhibitors against *M. tuberculosis* DNA gyrase enzyme ($MIC_{exp} \leq 1.0$ μg/mL), while the rest of the *ExpLib* molecules (31 compounds with $MIC_{exp} > 1.0$ μg/mL) belong to the class of inactive 6-FQs.

On the other hand, the *CombiLib* library comprised of 1.101 mixed "drug-like" 6-FQs was assembled *in-house*, by employing a classical virtual reaction-based combinatorial enumeration approach (Minovski & Šolmajer, 2010; Minovski et al., 2012). At the beginning, the original synthetic pathways of three most frequently used 6-FQs in tuberculosis chemotherapy, i.e., ciprofloxacin (CIP), moxifloxacin (MOX), and ofloxacin (OFL), were exploited as templates for construction of their virtual combinatorial synthetic schemes (Schwalbe et al., 2000; Martel et al., 1997; Serradel et al., 1983), while a commercial pre-filtered "*Ro3*" fragments library comprised of 6.995 building-blocks (Key Organics Bionet Fragment Library "Rule of 3"[19]) was utilized as a reactants subset for performing SAR-based fragmental permutations at the pre-defined 6-FQs variable scaffold positions (Figure 6b) for each virtual synthetic pathway separately (R_1 and R_7 substructural modifications for CIP and MOX, while only R_7 substructural modifications for OFL analogs). According to the current SAR knowledge of 6-FQs (Figure 6b), the position R_1 of the CIP and MOX scaffold was targeted by primary amines (e.g., R_1-NH_2), while the position R_7 in all three cases (CIP, MOX, and OFL) was targeted by secondary amines (e.g., R_1, R_2-NH). Moreover, although

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 2. Virtual combinatorial definitions and substructural fragments used for construction of six different virtual combinatorial libraries of 6-FQ structural analogs

ID	6-FQs	Virtual Combinatorial Definition		Substructural Fragments		$\rho_{tot} = N_i \times M_j$
		R_1	R_7	R_1 (N_i)	R_7 (M_j)	
1.	CIP-N_i-M_j	R_1-NH_2	R_2, R_3-NH	116	106	12.296
2.	CIP'-N_i-M_j	R_1-NH_2	R_2-*	115	191	21.965
3.	MOX-N_i-M_j	R_1-NH_2	R_2, R_3-NH	115	74	8.510
4.	MOX'-N_i-M_j	R_1-NH_2	R_2-*	73	147	10.731
5.	OFL-M_j	N/A	R_1, R_2-NH	N/A	180	180
6.	OFL'-M_j	N/A	R_1-*	N/A	189	189
Sum						**53.871**

(Minovski et al., 2012).

contrary to the current SAR recommendations, an additional structural diversity was introduced by attaching non-amino fragments (e.g., R-*) at the position R_7 in all three cases (CIP, MOX, and OFL). This resulted in obtaining a total of six reaction-based enumeration definitions (for more information check Minovski et al., 2012) for generation of six different 6-FQs virtual combinatorial libraries (Table 2).

As demonstrated in Table 2, the performed virtual reaction-based combinatorial enumerations yielded a general virtual combinatorial library comprised of total 53.871 compounds (CIP, MOX, and OFL structural analogs). Although, the resulting library could generally be considered as a small sub-space within the available chemical space (Bohacek et al., 1996), one should be aware about the very low probability that each compound in such a virtual compound library possess desired "drug-like" properties (Walters et al., 1999; Hann et al., 2001; Muegge et al., 2003). Therefore, in order to distill only those combinatorially generated 6-FQs delineating the "drug-like" chemical sub-space, the entire virtual combinatorial library was subjected to a robust druggability properties assessment. Assuming that our generated 6-FQs would be administered *per os*, we constructed a combined Lipinski-Veber "drug-likeness" filtering tool for automated isolation of the desired sub-space of "drug-like" 6-FQs (Minovski et al., 2012). This filtering strategy produced a list of 1.101 "drug-like" virtual combinatorial 6-FQ analogs, which were further used as an external dataset for prediction of their unknown biological activity values ($MIC_{pred\text{-}combi}$ [μg/mL]) by employing a pre-constructed predictive QSAR model (Minovski et al., 2011, Minovski et al. 2012).

Development of an ANNs-Based Predictive QSAR Model

Following the integrated *in silico* screening protocol described previously (Figure 2), non-linear artificial neural networks (ANNs)-based predictive QSAR modeling was performed on the *ExpLib* library of 145 6-FQs using a comprehensive set of approximately 600 calculated 2D molecular descriptors (Minovski et al., 2011). The heuristic algorithm that is a step-wise selection procedure was initially used for descriptors pool reduction to a sub-pool comprised of up to 10 significant parameters, while its further reduction was achieved by construction of an inter-correlation descriptor matrix and subsequent elimination of those descriptors for which the inter-correlation coefficient is $R^2(P_i, P_m) \leq 0.40$. Here, a proposed upper R^2 value of 0.40 was used in order to avoid a possible chance correlation during the modeling (Topliss, 1983). A total of 7 molecular descriptors retained at the end of the parameters reduction/selection pro-

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

cedure, which further served as independent input variables for the ANNs model development. Kohonen artificial neural networks (KANNs) were employed for division of the compound's data (Kohonen, 1982; Novič & Zupan, 1995; Zupan & Gasteiger, 1999) on a training set (115 compounds) and an external validation set (30 compounds), while counter-propagation artificial neural networks (CP ANNs) were used for the purpose of modeling (Zupan et al., 1995; Zupan et al., 1997; Zupan & Gasteiger, 1999).

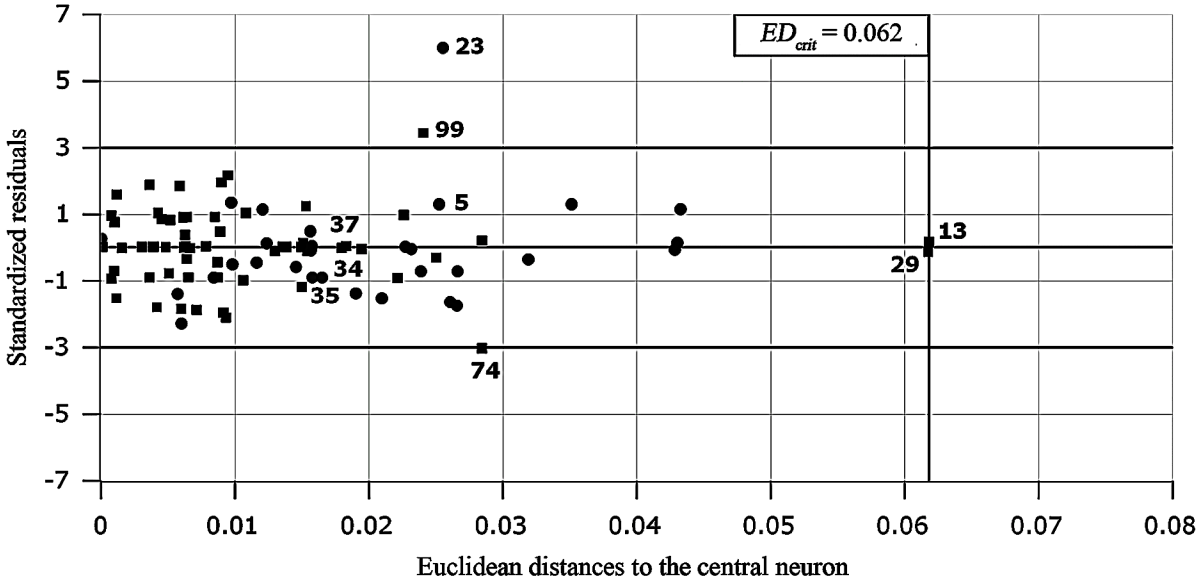
As stated previously, the modeling was performed solely on the training set objects, where various network architectures and number of learning epochs were extensively evaluated in the search for the most optimal CP ANN predictive model ($R_{tr} = 0.96$). The model was internally validated by using a CV LOO procedure ($R_{tr\text{-}cv} = 0.62$) as well as externally validated for its predictive performances using the previously excluded validation set objects ($Q_{ext} = 0.84$). Furthermore, the applicability domain (AD) of thus established and validated CP ANN predictive model was assessed by the *minimum Euclidean distance space* (MEDS) approach (Minovski, Župerl et al., 2013) – an efficient distance-based AD estimation method reflecting the reliability of the established CP ANN predictive model through utilization of the Euclidean distance (ED) metric in the model's structure-representation vector space (Figure 7).

As demonstrated in Figure 7, no external validation set compounds could be identified as outliers according to critical ED value ($ED_{crit} = 0.062$), i.e., a valuable information indicating that no significant structural differences exist between the investigated 6-FQs (structure-representation vector space). On the other hand, two training set objects (ID = 74 and 99) are somehow wrongly predicted by the model as demonstrated by their calculated standardized residual values, which are slightly above the $\pm 3\sigma$ boundaries. Among the external validation set objects, only one compound (ID = 23) could be determined as an extreme point with calculated standardized residual value around 6.0, i.e., a very poor biological activity prediction ($p\text{MIC}_{exp[\text{ID}=23]} = 0.7404$, $p\text{MIC}_{pred[\text{ID}=23]} = 0.2334$). The structural analysis of this compound shows that it belongs to a group of unconventional 6-FQs regarding the main scaffold (1,8-naphthyridine instead of quinoline moiety). In addition to this one, five more *ExpLib* compounds share the same structural feature, of which two (ID = 29 and 37) are training set objects, while three of them (ID = 5, 34, and 35) belong to the external validation set, and apparently all of them have acceptably predicted biological activity values, as they are situated within the boundaries of $\pm 3\sigma$ units (Figure 7). However, (ID = 23) is the only *ExpLib* compound that has attached (*S*)-3-aminopirrolydinil group at the position R_7 of the naphthyridine ring and no similar compounds exist in the training set. These findings explicitly pinpoint to a certain degree of weakness of the established CP ANN model to correctly predict the biological activity values for naphthyridine-like 6-FQs with (*S*)-3-aminopirrolydinil fragments attached at the R_7 position. Anyhow, the rest of the external validation set objects are situated within the boundaries of the model's AD – a result that clearly confirms the reliability of the constructed CP ANN predictive model for its further utilization as an efficient tool for evaluation of the biological activities for novel, not yet synthesized 6-FQs.

Finally, the established CP ANN predictive model was used for prediction of the biological activity values ($\text{MIC}_{pred\text{-}combi}$ [μg/mL]) for our previously generated 1.101 "drug-like" 6-FQ combinatorial analogs (Minovski et al., 2012). The predicted biological activities for these compounds were in the range between $0.0021 \leq \text{MIC}_{pred\text{-}combi}$ [μg/mL] ≤ 6.3726, i.e., a mix of highly active 6-FQs, but also totally inactive representatives according to the *in vitro* inhibition assays performed so far (Aubrey et al., 2006; Pantel et al., 2011; Pantel et al., 2012). Taking this information into account, we defined a so-called global hypothetical activity (GHA) range $0.00 \leq \text{MIC}_{pred\text{-}combi}$ [μg/mL] ≤ 1.00, which served as an efficient activity-based filtering tool for our 1.101 *CombiLib* compounds. This filtering procedure, resulted in selection of a subset of 427 "drug-like", but also hypothetically "active" 6-FQ combinatorial

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 7. Graphical representation of the CP ANN predictive model's applicability domain assessed by the MEDS-based AD estimation method; the AD boundaries are defined by the training set object (ID = 13) with maximal ED to the central neuron (ED_{crit} = 0.062) and ±3σ units for the calculated standardized residuals (predictability of the model), respectively. The training set objects (115 compounds) are represented as solid dark-gray rectangles, while the external validation set objects (30 compounds) as solid light-gray circles.



analogs, which were further used in structure-based calculations for final identification and selection of novel 6-FQs as potential *M. tuberculosis* DNA gyrase inhibitors.

Construction of *M. tuberculosis* DNA Gyrase Protein Homology Models for Molecular Docking Calculations

Although, various separate topoisomerase IIA subunits originating from different bacterial species are solved recently by X-ray crystallography (Laponogov et al., 2009; Laponogov et al., 2010; Wohlkonig et al., 2010; Bax et al., 2010; Tretter et al., 2010; Darmon et al., 2012; Bouige et al., 2013), unfortunately the entire crystal structure of the *M. tuberculosis* DNA gyrase holoenzyme in complex with the DNA molecule and an intercalated 6-FQ ligand still remains an indecipherable issue (Collin et al., 2011). Therefore, taking into consideration the recently proposed 6-FQs-topoisomerase binding mechanisms (Laponogov et al., 2010; Wohlkonig et al., 2010; Bax et al., 2010) as well as the structural and functional similarity between the both topoisomerase IIA paralogous forms (DNA gyrase and topoisomerase IV), we constructed three *M. tuberculosis* DNA gyrase protein homology models (Minovski et al., 2013).

The available topoisomerase IIA-DNA-6-FQ crystal structure complexes originating from three different bacterial species (*Streptococcus pneumoniae* topo IV-DNA-levofloxacin, PDB ID: 3K9F; *Acinetobacter baumannii* topo IV-DNA-moxifloxacin, PDB ID: 2XKK; and *Staphylococcus aureus* DNA gyrase-DNA-ciprofloxacin, PDB ID: 2XCT) and recently determined crystal structures of the separate *M. tuberculosis* DNA gyrase subunits - GyrA breakage-reunion domain, GyrA-BRD (PDB ID: 3IFZ)

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

and GyrB-Toprim domain (PDB ID: 3M4I) were utilized for protein sequence alignment, while the proposed key amino acid sequences covering the QBP in *M. tuberculosis* DNA gyrase (Piton et al., 2010) were exploited for structural interchange of the original QBP sequences. It should be stressed that during the modeling, the nascent conformations of the DNA molecule, the intercalated ligands (levofloxacin [LFX], moxifloxacin [MOX], and ciprofloxacin [CIP], respectively), as well as the contributing cofactors present in the QBPs (e.g., water molecule(s) and/or Mg^{2+} ion(s)) were left intact. Moreover, the experimentally determined spatial coordinates of the co-crystallized 6-FQ conformations (LFX, MOX, and CIP, respectively) were used to define the QBP (a screening area with cavity radius of 12.5 Å) for each constructed homology model (for more details check Minovski et al., 2013).

The *M. tuberculosis* DNA gyrase protein homology models thus assembled and prepared (named as $3K9F_{mod}$, $2XKK_{mod}$, and $2XCT_{mod}$, respectively) which QBP emulates the one present in the wild-type enzyme (PDB IDs: 3IFZ and 3M4I), were further used as starting points to perform molecular docking calculations.

Molecular Docking Calculations and Protein Homology Models Validation

The molecular docking calculations on both 6-FQ compound libraries (*ExpLib* and *CombiLib*), within the previously defined QBP of each *M. tuberculosis* DNA gyrase protein homology model ($3K9F_{mod}$, $2XKK_{mod}$, and $2XCT_{mod}$, respectively), were performed by using the GOLD docking suite[20] (Jones et al., 1997). The entire docking procedure (including all the GA settings and required technical parameters) for each protein homology model was accomplished as described in our previous work (Minovski et al., 2013), while the GOLDScore Fitness (GSF) function was used for estimation of the 6-FQs binding affinity.

At the beginning, the quality of the constructed protein homology models was confirmed by an initial re-docking validation run (ligand reproduction assessment) on the co-crystallized ligand conformations (LFX, MOX, and CIP, respectively) present within the QBP of each protein homology model separately, followed by heavy-atoms RMSD (Å) comparison between the experimental ligand conformations and their corresponding GOLD-derived ligand poses (Figure 8a and Table 3).

As demonstrated in Table 3, a total of three dock poses were computed for each co-crystallized ligand. Except $2XKK_{mod}$ model for which the initial validation failed according to the calculated RMSD (Å) values for MOX-reproduced dock poses (RMSD > 2.0 Å; an apparently poor model), the rest two homology models ($3K9F_{mod}$ and $2XCT_{mod}$) successfully reproduced their experimental ligand conformations (LFX and CIP, respectively) with calculated RMSD values significantly below 2.0 Å (Verdonk et al., 2003). These preliminary validation results suggest the following order of quality of our constructed *M. tuberculosis* DNA gyrase protein homology models ($2XKK_{mod} < 2XCT_{mod} < 3K9F_{mod}$), which result could also be visually determined (Figure 8a).

However, to thoroughly assess their quality as well as reliability as efficient VS filters, a second validation experiment of their discriminatory performances (an investigation of the models capability to correctly discriminate between known active and inactive compounds) was carried out as described previously (Hevener et al., 2009). For that purposes, the *ExpLib* library of 145 6-FQs was first docked within the QBP of each protein homology model ($3K9F_{mod}$, $2XKK_{mod}$, and $2XCT_{mod}$, respectively) by using the same GA settings, and afterwards only the subset of 114 active top-scored dock poses ($MIC_{exp} \leq 1.0$ μg/mL) was enriched with a multiconformer library of total 13.990 artificial decoy molecules randomly selected from the Asinex Elite Library[21] (Minovski et al., 2013). Thus prepared, the library of active/decoy molecules was subjected to a robust similarity screening against the co-crystallized ligand

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 8. Validation of the constructed M. tuberculosis DNA gyrase protein homology models. (a) Re-docking of the co-crystallized ligand conformations (LFX, MOX, and CIP, respectively) within the QBP of each protein homology model (3K9F$_{mod}$, 2XKK$_{mod}$, and 2XCT$_{mod}$), separately. The experimental ligand conformations and the corresponding GOLD-calculated ligand poses are depicted in dark and light gray, respectively (stick representation), the water molecule(s) are represented as black sphere(s), while Mg^{2+} ion as a light gray sphere. (b) Protein homology models discriminatory performances for correct identification of active and inactive 6-FQs assessed by ROC methodology. The diagonal line (line of no discrimination) corresponds to the randomly distributed data (RDD = 0.5).

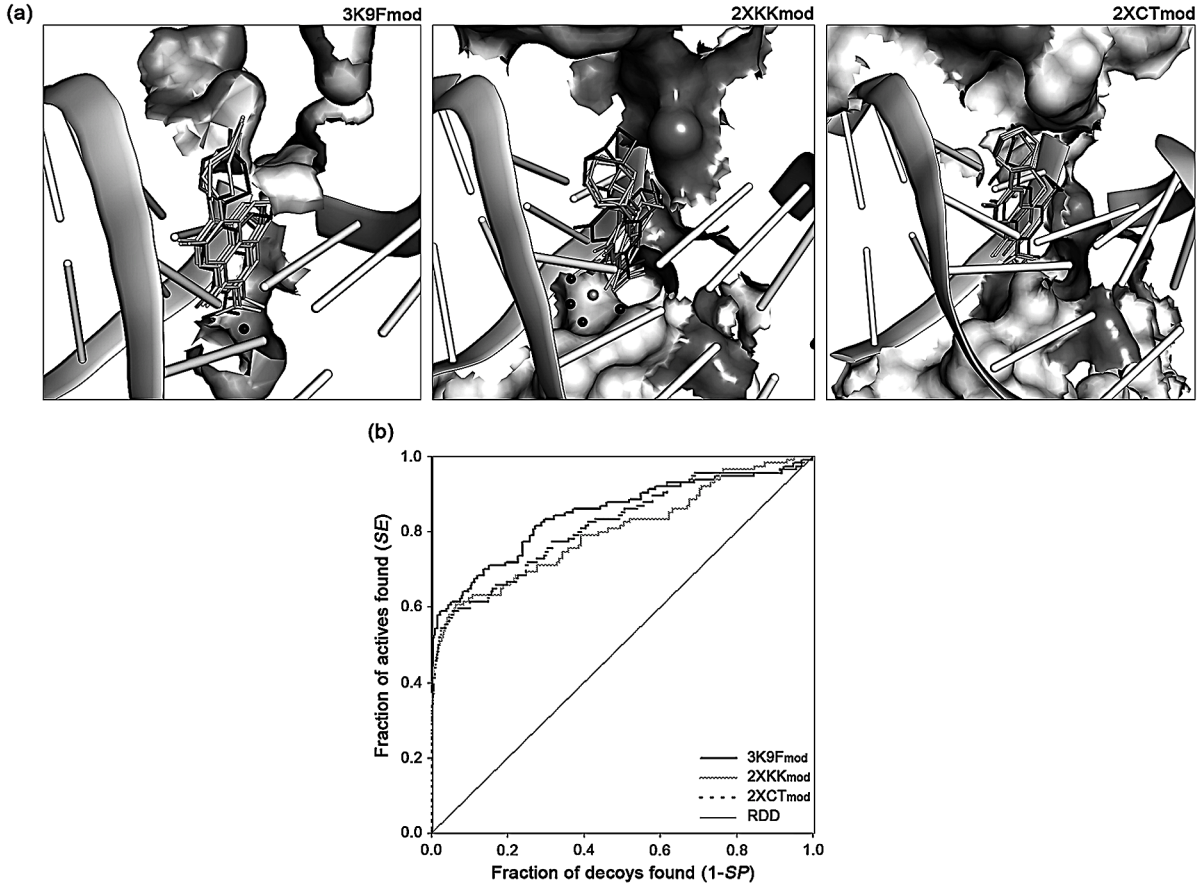


Table 3. Ligand reproduction assessment for initial validation of the quality of the assembled M. tuberculosis DNA gyrase protein homology models (3K9F$_{mod}$, 2XKK$_{mod}$, and 2XCT$_{mod}$, respectively) performed by heavy-atoms RMSD (Å) comparison between each experimental ligand conformation (LFX, MOX, and CIP, separately) and their corresponding GOLD-computed dock poses

Model	**3K9F$_{mod}$ [LFX]**			**2XKK$_{mod}$ [MOX]**			**2XCT$_{mod}$ [CIP]**		
Dock pose	pose 1	pose 2	pose 3	pose 1	pose 2	pose 3	pose 1	pose 2	pose 3
RMSD (Å)	1.056	1.121	1.027	2.553	2.652	2.488	1.249	1.311	1.340

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 4. The statistical parameters describing the VS performances of our constructed M. tuberculosis DNA gyrase protein homology models, calculated directly from the obtained ROC curves: ROC-AUC (area under the ROC curve), EF (enrichment factor), and p-values (the probability for a model to retrieve better outcome then the reference one, assuming that the null hypothesis is true). All comparisons are performed relative to the 3K9F$_{mod}$ model used as reference.

Model	3K9F$_{mod}$	2XKK$_{mod}$	*p*-Value	2XCT$_{mod}$	*p*-value
ROC-AUC	0.844 [0.795, 0.891]	0.815 [0.769, 0.864]	0.192	0.826 [0.774, 0.870]	0.299
EF (0.5%)	85.464 [66.67, 105.60]	68.829 [50.88, 89.23]	0.108	77.032 [57.14, 97.48]	0.273
EF (1.0%)	52.046 [45.50, 61.46]	43.473 [34.62, 52.73]	0.091	46.779 [36.84, 56.67]	0.215
EF (2.0%)	27.537 [22.80, 32.66]	24.786 [20.08, 29.31]	0.203	27.365 [23.61, 32.00]	0.479

conformations (LFX, MOX, and CIP, respectively) used as queries (Huang et al., 2006). vROCS tool[22] was used to evaluate the VS discriminatory performances of the constructed protein homology models (ROC-AUC and early enrichment parameters at 0.5, 1.0, and 2.0% retrieved within ±95% confidence interval) directly from the generated ROC curves (Figure 8b and Table 4).

Table 4 summarizes the statistical parameters carrying the discriminatory performances of our constructed *M. tuberculosis* DNA gyrase protein homology models as efficient VS filters calculated directly from the generated ROC curves where all the comparisons are performed relative to the 3K9F$_{mod}$ model (Figure 8b). As displayed here, according to the highest calculated ROC-AUC value the best VS performances could be assigned to 3K9F$_{mod}$ model (ROC-AUC$_{[3K9Fmod]} = 0.844$), 2XCT$_{mod}$ has somehow moderate VS performances (ROC-AUC$_{[2XCTmod]} = 0.826$), while 2XKK$_{mod}$ could be identified as a poorest one (ROC-AUC$_{[2XKKmod]} = 0.815$). These findings are also supported by the calculated *p*-values – a probability for an investigated model to give a better outcome then the reference one (in our case 3K9F$_{mod}$) under assumption that the null hypothesis is true. As demonstrated, the calculated ROC-AUC *p*-values for 2XKK$_{mod}$ and 2XCT$_{mod}$ models both tend toward 0 (p-value$_{[2XKKmod]} = 0.192$ and p-value$_{[2XCTmod]} = 0.299$, respectively), i.e., the probability for these two models to give a better outcome (identification of more active molecules than the reference model) is practically very low and this result is not due to a random chance (p-value < 0.5; the null hypothesis could be rejected). Similarly, the probability for 2XKK$_{mod}$ to give better outcome than 2XCT$_{mod}$ is very low as well (p-value$_{[2XCTmod]} > p$-value$_{[2XKKmod]}$), which results are additionally grounded by the estimated early enrichment factors (calculated at 0.5, 1.0, and 2.0% in the boundaries of ±95% confidence interval) and their corresponding probability values. Moreover, these results are also congruent with the initial re-docking validation results (Figure 8a and Table 3), allowing us to confirm our previously established quality order of our protein homology models (2XKK$_{mod}$ < 2XCT$_{mod}$ < 3K9F$_{mod}$). Based on both validation experiments, the 3K9F$_{mod}$ could be recognized as a model with by far the best VS discriminatory performances, and therefore it was selected for further utilization as a robust *in silico* filtering device for VS of our novel 6-FQ combinatorial analogs.

Finally, the *CombiLib* library of 427 "drug-like" 6-FQ combinatorial analogs was docked into QBP of the selected *M. tuberculosis* DNA gyrase protein homology model (3K9F$_{mod}$), while the experimental conformation of the co-crystallized LFX ligand present within the QBP was used as a template structure against which an *in-depth* post-docking analysis of the GOLD-computed docking solutions was performed (Minovski et al., 2013).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Boolean-Based Clustering for Identification and Selection of Novel 6-FQ Hits

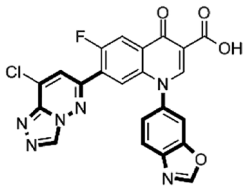
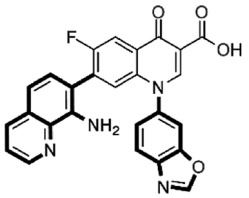
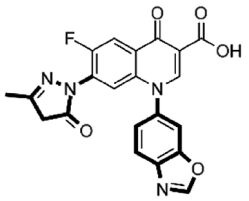
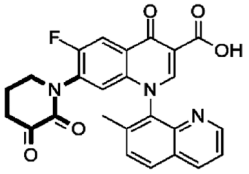
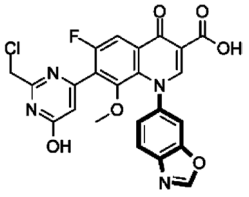
As described previously, the generated docking solutions are usually stored in a form of extended virtual compound libraries where all computed dock poses per ligand are organized as small clusters. Therefore, to thoroughly assess all the generated docking solutions for our 427 "drug-like" *CombiLib* 6-FQ analogs (a total of 1281 poses arranged into 427 clusters, i.e., 3 poses per cluster) as well as to identify and select the most promising 6-FQ hit candidates, each calculated dock pose was subjected to a detailed post-docking analysis based on their visual examination using the available structural data.

Following the Boolean-based (T/F) clustering method described in the previous section, the post-docking analysis was carried out in three consecutively-coupled levels: geometric properties assessment, score-based clustering, and activity-based clustering (Minovski et al., 2013). Within the scope of the first level of the performed post-docking analysis (geometric properties assessment), a total of 162 (T)-signed 6-FQ analogs were initially identified as geometrically well positioned relative to the experimental co-crystallized LFX conformation. The structural analysis of these compounds clearly reflected the capability of the selected protein homology model $3K9F_{mod}$ to correctly identify 6-FQ structural analogs that belong to different structural classes (CIP, MOX, and OFL *CombiLib* compounds). All (T)-signed 6-FQ combinatorial analogs thus selected were extracted as a separate cluster and used in the second level of the Boolean-based (T/F) post-docking analysis by taking into consideration their estimated GSF function (score-based clustering). Implementing a pre-defined GSF threshold of ($GSF \geq 80$), a total of 92 top-scored clusters were determined and accordingly a total of 92 (T)-signed top-scored 6-FQ combinatorial analogs were selected. Similarly, the selected docking solutions were distilled as a separate cluster and used as an input in the final level of the Boolean-based (T/F) post-docking analysis (activity-based clustering) by taking into account their QSAR-predicted biological activity values ($MIC_{pred\text{-}combi}$ [μg/mL]). It should be stressed once again that all 6-FQ analogs comprising our *CombiLib* library are somehow hypothetically active as inhibitors of *M. tuberculosis* DNA gyrase enzyme according to the previously implemented GHA range ($0.00 \leq MIC_{pred\text{-}combi}$ [μg/mL] ≤ 1.00). Nevertheless, in order to extract the most "active" 6-FQ hits, the cluster of (T)-signed highly-scored hits assembled at the end of the previous level, was subjected to an additional activity-based filtering routine ($MIC_{pred\text{-}combi} \leq 0.05$ μg/mL). In this procedure, a total of 48 (T)-signed combinatorial hits were identified as most "active", mainly CIP and MOX structural analogs of which the most promising are listed in Table 5, while no OFL combinatorial compounds were identified at the end of the Boolean-based (T/F) post-docking analysis.

The substructural examination of the building-blocks attached at the pre-defined variable scaffold positions of the isolated 6-FQ combinatorial hits revealed the most frequently occurring fragments (Table 5). Namely, at R_1 position benzo[*d*]oxazole was found as the most frequently occurring one (fragment **028** in CIP- and CIP'-analogs, which is relevant to the fragment **016** in MOX'-analogs). On the other hand, four fragments were identified as the most frequently appearing at R_7 position including 8-chloro-6-methyl-[1,2,4]triazolo[4,3-*b*]pyridazine (fragment **059**), 7-methylquinolone-8-amine (fragment **073**), 1-methylpiperidine-2,3-dione (fragment **096**), and 1,3-dimethyl-1*H*-pyrazol-5(4*H*)-one (fragment **102**). From a medicinal chemistry perspective, these fragments are mainly small aromatic N-heterocyclic systems with molecular weight between 98 and 155 g/mol, which contain more than two HBA atoms on average – an important SAR feature that undoubtedly increases the probability of establishing HB interactions between 6-FQs and surrounding amino acid residues of the GyrA/GyrB subunit of the enzyme.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 5. Some promising CombiLib 6-FQ representatives identified by using the Boolean-based (T/F) post-docking analysis; the most frequently occurring substructural fragments at R_1 and R_7 position are represented in bold.

ID	Code	*CombiLib* Hits	R_1	R_7	$MIC_{pred\text{-}combi}$ [μg/mL]	GSF
1.	CIP'-**028-059**		**028**	**059**	0.0499	83.55
2.	CIP'-**028-073**		**028**	**073**	0.0021	88.89
3.	CIP-**028-102**		**028**	**102**	0.013	80.87
4.	CIP-049-**096**		049	**096**	0.0329	86.57
5.	MOX'-**016**-137		**016**	137	0.03	89.89

CONCLUSION

In the current era of enormous technological advancements, we have witnessed a rapid breakthrough in the modern design and optimization of novel drug candidates. Nowadays, the *in silico* drug design methods are indeed an irreplaceable supplement to the experimentally grounded approaches not only in the early hit(s) identification and hit-to-lead stages, but also in the late points of the drug discovery pipeline. From the profusion of *in silico* drug discovery methods currently available, various two- and three-dimensional approaches (e.g., QSAR, ligand-based, structure-based, etc.) became a cornerstone of the modern drug discovery for rapid and efficient development of novel successful drug candidates.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Unfortunately, it was found that their individual utilization as *in silico* ligand filters could frequently result in a very small number of newly identified or *de novo* developed drug candidates, and therefore, more and more efforts are currently devoted to a skillful, rational, and knowledge-based development of integrated *in silico* drug discovery platforms.

In this chapter, we illustrated such an integrated *in silico* screening platform for fast and efficient identification of novel drug candidates from a large number of possibilities. A variety of well-established *in silico* drug discovery methods were covered and their specific assemblage into an efficient screening integration for identification and selection of novel hit candidates was thoroughly reviewed. Finally, we illustrated its practical application in a case study related to the design and identification of novel 6-FQ antibacterials as potential *M. tuberculosis* DNA gyrase inhibitors and proposed some new SAR guidelines. In conclusion, we believe that the future of the modern drug development resides in the implementation of such integrated *in silico* screening schemes, which could aid not only the development of novel and potent drugs, but could also introduce some novel standards in the on-going drug discovery programs.

REFERENCES

Abuhamdah, S., Habash, M., & Taha, M. O. (2013). Elaborate ligand-based modeling coupled with QSAR analysis and *in silico* screening reveal new potent acetylcholinesterase inhibitors. *Journal of Computer-Aided Molecular Design*, *27*(12), 1075–1092. doi:10.1007/s10822-013-9699-6 PMID:24338032

Agrafiotis, D. K. (2000). Multiobjective optimization of combinatorial libraries. *Molecular Diversity*, *5*(4), 209–230. doi:10.1023/A:1021320124615 PMID:12549673

Agrafiotis, D. K., Lobanov, V. S., & Salemme, F. R. (2002). Combinatorial informatics in the post-genomic era. *Nature Reviews. Drug Discovery*, *1*(5), 337–346. doi:10.1038/nrd791 PMID:12120409

Anderson, A. C. (2003). The process of structure-based drug design. *Chemistry & Biology*, *10*(9), 787–797. doi:10.1016/j.chembiol.2003.09.002 PMID:14522049

Andersson, P. M., Sjöström, M., Wold, S., & Lundstedt, T. (2001). Strategies for subset selection of parts of an in-house chemical library. *Journal of Chemometrics*, *15*(4), 353–369. doi:10.1002/cem.671

Aronov, A. M. (2002). Design of virtual combinatorial libraries. In L. Bellavance English (Ed.), *Methods in molecular biology, combinatorial library: Methods and protocols* (Vol. 201, pp. 267–276). Totowa, NJ: Humana Press; doi:10.1385/1-59259-285-6:267

Aubrey, A., Veziris, N., Cambau, E., Truffot-Pernot, C., Jarlier, V., & Fisher, L. M. (2006). Novel gyrase mutations in quinolone-resistant and –hypersusceptible clinical isolates of *Mycobacterium tuberculosis*: Functional analysis of mutant enzymes. *Antimicrobial Agents and Chemotherapy*, *50*(1), 104–112. doi:10.1128/AAC.50.1.104-112.2006 PMID:16377674

Bajorath, J. (2002). Integration of virtual and high-throughput screening. *Nature Reviews. Drug Discovery*, *1*(11), 882–894. doi:10.1038/nrd941 PMID:12415248

Bajot, F. (2006). The use of QSAR and computational methods in drug design. In T. Puzyn, J. Leszczynski, & M. T. D. Cronin (Eds.), *Recent advances in QSAR studies: Methods and applications* (pp. 261–282). Springer Dordreht.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Barril, X. (2012). Druggability predictions: Methods, limitations, and applications. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, *3*(4), 327–338. doi:10.1002/wcms.1134

Bartfai, T., & Lees, G. V. (2006). The business basics (general). In T. Bartfai & G. V. Lees (Eds.), *Drug discovery: From bedside to Wall Street* (pp. 183–192). Oxford, UK: Elsevier Academic Press.

Bax, B. D., Chan, P. F., Eggleston, D. S., Fosberry, A., Gentry, D. R., & Gorec, F. et al. (2010). Type IIA topoisomerase inhibition by a new class of antibacterial agents. *Nature*, *466*(7309), 935–940. doi:10.1038/nature09197 PMID:20686482

Bhal, S. K., Kassam, K., Peirson, I. G., & Pearl, G. M. (2007). The rule of five revisited: Applying logD in place of logP in drug-likeness filters. *Molecular Pharmaceutics*, *4*(4), 556–560. doi:10.1021/mp0700209 PMID:17530776

Bharath, E. N., Manjula, S. N., & Vijaychand, A. (2011). *In Silico* drug design – Tool for overcoming the innovation deficit in the drug discovery process. *International Journal of Pharmacy and Pharmaceutical Sciences*, *3*(2), 8–12.

Bleicher, K. H., Böhm, H.-J., Müller, K., & Alanine, A. I. (2003). Hit and lead generation: Beyond high-throughput screening. *Nature Reviews. Drug Discovery*, *2*(5), 369–378. doi:10.1038/nrd1086 PMID:12750740

Bohacek, R. S., mcMartin, C., & Guida, W. C. (1996). The art and practice of structure-based drug design. *Medicinal Research Reviews*, *16*(1), 3–50. doi:10.1002/(SICI)1098-1128(199601)16:1<3::AID-MED1>3.0.CO;2-6 PMID:8788213

Borchardt, R. T. (2004). Scientific, educational and communication issues associated with integrating and applying drug-like properties in drug discovery. In R. Borchardt, E. Kerns, C. Lipinski, D. Thakker, & B. Wang (Eds.), *Pharmacological profiling in drug discovery for lead selection* (pp. 451–466). Arlington, VA: AAPS Press.

Bouige, A., Darmon, A., Piton, J., Roue, M., Petrella, S., & Capton, E. et al. (2013). *Mycobacterium tuberculosis* DNA gyrase possesses two functional GyrA-boxes. *The Biochemical Journal*, *455*(3), 285–294. doi:10.1042/BJ20130430 PMID:23869946

Bouvier, G., Evrard-Todeschi, N., Girault, J.-P., & Bertho, G. (2010). Automatic clustering of docking poses in virtual screening process using self-organizing map. *Bioinformatics (Oxford, England)*, *26*(1), 53–60. doi:10.1093/bioinformatics/btp623 PMID:19910307

Burden, F. R., & Winkler, D. A. (1999). New QSAR method applied to structure-activity mapping and combinatorial chemistry. *Journal of Chemical Information and Computer Sciences*, *39*(2), 236–242. doi:10.1021/ci980070d

Cavazzani, P., & Rajagopal, N. A. (2010). Market analysis of growing competition in pharmaceutical industry. *International Journal of Business Competition and Growth*, *1*(1), 31–44. doi:10.1504/IJBCG.2010.032827

Chabala, J. C. (1995). Solid-phase combinatorial chemistry and novel tagging methods for identifying leads. *Current Opinion in Biotechnology*, *6*(6), 632–639. doi:10.1016/0958-1669(95)80104-9 PMID:8527833

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chen, C. Y. (2013). A novel integrated framework and improved methodology of computer-aided drug design. *Current Topics in Medicinal Chemistry*, *13*(9), 965–988. doi:10.2174/1568026611313090002 PMID:23651478

Cheng, T., Li, Q., Zhou, Z., Wang, Y., & Bryant, S. H. (2012). Structure-based virtual screening for drug discovery: A problem-centric review. *The AAPS Journal*, *14*(1), 133–141. doi:10.1208/s12248-012-9322-0 PMID:22281989

Chirico, N., & Gramatica, P. (2011). Real external predictivity of QSAR models: How to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient. *Journal of Chemical Information and Modeling*, *51*(9), 2320–2335. doi:10.1021/ci200211n PMID:21800825

Clark, D. E. (2003). In silico prediction of blood-brain barrier permeation. *Drug Discovery Today*, *8*(20), 927–933. doi:10.1016/S1359-6446(03)02827-7 PMID:14554156

Coe, D. M., & Storer, R. (1999). Solution-phase combinatorial chemistry. *Molecular Diversity*, *4*(1), 31–38. doi:10.1023/A:1009694409264 PMID:10320987

Coi, A., & Bianucci, A. M. (2013). Combining structure- and ligand-based approaches for studies of interactions between different conformations of hERG K^+ channel pore and known ligands. *Journal of Molecular Graphics & Modelling*, *46*, 93–104. doi:10.1016/j.jmgm.2013.10.001 PMID:24185260

Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., & Harris, D. et al. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, *393*(6685), 537–544. doi:10.1038/31159 PMID:9634230

Collin, F., Karkare, S., & Maxwell, A. (2011). Exploiting bacterial DNA gyrase as a drug target: Current state and perspectives. *Applied Microbiology and Biotechnology*, *92*(3), 479–497. doi:10.1007/s00253-011-3557-z PMID:21904817

Congreve, M., Carr, R., Murray, C. W., & Jhoti, H. (2003). A "rule of three" for fragment-based lead discovery. *Drug Discovery Today*, *8*(19), 876–877. doi:10.1016/S1359-6446(03)02831-9 PMID:14554012

Consonni, V., Ballabio, D., & Todeschini, R. (2009). Comments on the definition of the Q^2 parameter for QSAR validation. *Journal of Chemical Information and Modeling*, *49*(7), 1669–1678. doi:10.1021/ci900115y PMID:19527034

Consonni, V., Ballabio, D., & Todeschini, R. (2010). Evaluation of model predictive ability by external validation techniques. *Journal of Chemometrics*, *24*(3-4), 194–201. doi:10.1002/cem.1290

Cramer, R. D., Patterson, D. E., & Bunce, J. D. (1988). Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *Journal of the American Chemical Society*, *110*(18), 5959–5967. doi:10.1021/ja00226a005 PMID:22148765

Cruz-Monteagudo, M., Borges, F., Cordeiro, M. N. D. S., Fajin, J. L. C., Morell, C., & Ruiz, R. M. et al. (2008). Desirability-based methods of multiobjective optimization and ranking for global QSAR studies. Filtering safe and potent drug candidates from combinatorial libraries. *Journal of Combinatorial Chemistry*, *10*(6), 897–913. doi:10.1021/cc800115y PMID:18855460

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Cumming, J. G., Davis, A. M., Muresan, S., Haeberlein, M., & Chen, H. (2013). Chemical predictive modeling to improve compound quality. *Nature Reviews. Drug Discovery*, *12*(12), 948–962. doi:10.1038/nrd4128 PMID:24287782

Darmon, A., Piton, J., Roué, M., Petrella, S., Aubrey, A., & Mayer, C. (2012). Purification, crystallization and preliminary X-ray crystallographic studies of the *Mycobacterium tuberculosis* DNA gyrase CTD. *Acta Crystallographica. Section F, Structural Biology and Crystallization Communications*, *68*(Pt 2), 178–180. doi:10.1107/S1744309111051888 PMID:22297993

Deng, Z.-L., Du, C.-X., Li, X., Hu, B., Kuang, Z.-K., & Wang, R. et al. (2013). Exploring the biologically relevant chemical space for drug discovery. *Journal of Chemical Information and Modeling*, *53*(11), 2820–2828. doi:10.1021/ci400432a PMID:24125686

Doman, T. N., McGovern, S. L., Witherbee, B. J., Kasten, T. P., Kurumbail, R., & Stallings, W. C. et al. (2002). Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. *Journal of Medicinal Chemistry*, *45*(11), 2213–2221. doi:10.1021/jm010548w PMID:12014959

Drlica, K., & Malik, M. (2003). Fluoroquinolones: Action and resistance. *Current Topics in Medicinal Chemistry*, *3*(3), 249–282. doi:10.2174/1568026033452537 PMID:12570763

Drwal, M. N., & Griffith, R. (2013). Combination of ligand- and structure-based methods in virtual screening. *Drug Discovery Today. Technologies*, *10*(3), e395–e401. doi:10.1016/j.ddtec.2013.02.002 PMID:24050136

Ekins, S., Crumb, W. J., Sarazan, R. D., Wikel, J. H., & Wrighton, S. A. (2002). Three-dimensional quantitative structure-activity relationship for inhibition of human ether-a-go-go-related gene potassium channel. *The Journal of Pharmacology and Experimental Therapeutics*, *301*(2), 427–434. doi:10.1124/jpet.301.2.427 PMID:11961040

Ekins, S., Mestres, J., & Testa, B. (2007). *In silico* pharmacology for drug discovery: Methods for virtual ligand screening and profiling. *British Journal of Pharmacology*, *152*(1), 9–20. doi:10.1038/sj.bjp.0707305 PMID:17549047

Ellingson, S. R., & Baudry, J. (2012). High-throughput virtual molecular docking with AutoDockCloud. *Concurrency and Computation*, *26*(4), 907–916. doi:10.1002/cpe.2926

Ellingson, S. R., Dakshanamurthy, S., Brown, M., Smith, J. C., & Baudry, J. (2013). Accelerating virtual high-throughput ligand docking: Current technology and case study on a petascale supercomputer. *Concurrency and Computation*. doi:10.1002/cpe PMID:24729746

Ellingson, S. R., Smith, J. C., & Baudry, J. (2013). VinaMPI: Facilitating multiple receptor high-throughput virtual docking on high-performance computers. *Journal of Computational Chemistry*, *34*(25), 2212–2221. doi:10.1002/jcc.23367 PMID:23813626

Eriksson, L., Jaworska, J., Worth, A. P., Cronin, M. T. D., McDowell, R. M., & Gramatica, P. (2003). Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. *Environmental Health Perspectives*, *111*(10), 1361–1375. doi:10.1289/ehp.5758 PMID:12896860

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Esposito, E. X., Hopfinger, A. J., & Madura, J. D. (2004). Methods for applying the quantitative structure-activity relationship paradigm. In J. Bajorath (Ed.), *Methods in molecular biology, chemoinformatics: Concepts, methods, and tools for drug discovery* (Vol. 275, pp. 131–213). Totowa, NJ: Humana Press; doi:10.1385/1-59259-802-1:131

Ferrero, L., Cameron, B., Manse, B., Lagneaux, D., Crouzet, J., Famechon, A., & Blanche, F. (1994). Cloning and primary structure of Staphylococcus aureus DNA topoisomerase IV: A primary target of fluoroquinolones. *Molecular Microbiology*, *13*(4), 641–653. doi:10.1111/j.1365-2958.1994.tb00458.x PMID:7997176

Free, S. J., & Wilson, J. (1964). A mathematical contribution to structure-activity studies. *Journal of Medicinal Chemistry*, *7*(4), 395–399. doi:10.1021/jm00334a001 PMID:14221113

Gedeck, P., & Lewis, R. A. (2008). Exploiting QSAR models in lead optimization. *Current Opinion in Drug Discovery & Development*, *11*(4), 569–575. PMID:18600573

Gellert, M., Mizuuchi, K., O'Dea, M. H., Itoh, T., & Tomizawa, J.-I. (1977). Nalidixic acid resistance: A second genetic character involved in DNA gyrase activity. *Proceedings of the National Academy of Sciences of the United States of America*, *74*(11), 4772–4776. doi:10.1073/pnas.74.11.4772 PMID:337300

Ghosh, S., Nie, A., An, J., & Huang, Z. (2006). Structure-based virtual screening of chemical libraries for drug discovery. *Current Opinion in Chemical Biology*, *10*(3), 194–202. doi:10.1016/j.cbpa.2006.04.002 PMID:16675286

Gohlke, H., & Klebe, G. (2001). Statistical potentials and scoring functions applied to protein-ligand binding. *Current Opinion in Structural Biology*, *11*(2), 231–235. doi:10.1016/S0959-440X(00)00195-0 PMID:11297933

Golbraikh, A., & Tropsha, A. (2002). Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. *Journal of Computer-Aided Molecular Design*, *16*(5-6), 357–369. doi:10.1023/A:1020869118689 PMID:12489684

Gozalbes, R., Mosulén, S., Ortí, L., Rodríguez-Díaz, J., Carbajo, R. J., Melnyk, P., & Pineda-Lucena, A. (2013). Hit identification of novel heparanase inhibitors by structure- and ligand-based approaches. *Bioorganic & Medicinal Chemistry*, *21*(7), 1944–1951. doi:10.1016/j.bmc.2013.01.033 PMID:23415087

Gray, J. J., Moughon, S., Wang, C., Schueler-Furman, O., Kuhlman, B., Rohl, C. A., & Baker, D. (2003). Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *Journal of Molecular Biology*, *331*(1), 281–299. doi:10.1016/S0022-2836(03)00670-3 PMID:12875852

Gribbon, P., & Sewing, A. (2005). High-throughput drug discovery: What can we expect from HTS? *Drug Discovery Today*, *10*(1), 17–22. doi:10.1016/S1359-6446(04)03275-1 PMID:15676295

Groll, A. V., Martin, A., Jureen, P., Hoffner, S., Vandamme, P., & Portaels, F. et al. (2009). Fluoroquinolone resistance in *Mycobacterium tuberculosis* and mutations on *gyrA* and *gyrB*. *Antimicrobial Agents and Chemotherapy*, *53*(10), 4498–4500. doi:10.1128/AAC.00287-09 PMID:19687244

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Grzybowski, B. A., Ishchenko, A. V., Kim, C.-Y., Topalov, G., Chapman, R., & Christianson, D. W. et al. (2002). Combinatorial computational method gives new picomolar ligands for a known enzyme. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(3), 1270–1273. doi:10.1073/pnas.032673399 PMID:11818565

Gussio, P., Pattabiraman, N., Zaharevitz, D. W., Kellogg, G. E., Topol, I. A., & Rice, W. G. et al. (1996). All-atom models for the non-nucleoside binding site of HIV-1 reverse transcriptase complexed with inhibitors: A 3D QSAR approach. *Journal of Medicinal Chemistry*, *39*(8), 1645–1650. doi:10.1021/jm9508088 PMID:8648604

Halperin, I., Ma, B., Wolfson, H., & Nussinov, R. (2002). Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins: Structure, Function, and Bioinformatics*, *47*(4), 409–443. doi:10.1002/prot.10115 PMID:12001221

Hann, M. M., Leach, A. R., & Harper, G. (2001). Molecular complexity and its impact on the probability of finding leads for drug discovery. *Journal of Chemical Information and Computer Sciences*, *41*(3), 856–864. doi:10.1021/ci000403i PMID:11410068

Hann, M. M., & Oprea, T. I. (2004). Pursuing the leadlikeness concept in pharmaceutical research. *Current Opinion in Chemical Biology*, *8*(3), 255–263. doi:10.1016/j.cbpa.2004.04.003 PMID:15183323

Hansch, C. J., & Fujita, T. (1964). ρ-σ-π Analysis. A method for the correlation of biological activity and chemical structure. *Journal of the American Chemical Society*, *86*(8), 1616–1626. doi:10.1021/ja01062a035

Harris, C. J., Hill, R. D., Sheppard, D. W., Slater, M. J., & Stouten, P. F. W. (2011). The design and application of target-focused compound libraries. *Combinatorial Chemistry & High Throughput Screening*, *14*(6), 521–531. doi:10.2174/138620711795767802 PMID:21521154

Hecker, E. A., Duraiswami, C., Andrea, T. A., & Diller, D. J. (2002). Use of ctalyst pharmacophore models for screening of large combinatorial libraries. *Journal of Chemical Information and Modeling*, *42*(5), 1204–1211. doi:10.1021/ci020368a PMID:12377010

Hevener, K. E., Zhao, W., Ball, D. M., Babaoglu, K., Qi, J., White, S. W., & Lee, R. E. (2009). Validation of molecular docking programs for virtual screening against dihydropteroate synthase. *Journal of Chemical Information and Modeling*, *49*(2), 444–460. doi:10.1021/ci800293n PMID:19434845

Hopkins, A. L. (2008). Pharmacological space. In C. G. Wermuth (Ed.), *The practice of medicinal chemistry* (3rd ed.; pp. 521–532). Amsterdam, Netherlands: Academic Press/Elsevier. doi:10.1016/B978-0-12-374194-3.00025-1

Hsu, C. –H., Lin, C. –Y., Ouyang, M., & Guo, Y. K. (2013). Biocloud: Cloud computing for biological, genomics, and drug design. *BioMed Research International, 2013*(Article ID 909470), 1-3. doi: 10.1155/2013/909470

Hu, S., Yu, H., Zhao, L., Liang, A., Liu, Y., & Zhang, H. (2013). Molecular docking and 3D-QSAR studies on checkpoint kinase 1 inhibitors. *Medicinal Chemistry Research*, *22*(10), 4992–5013. doi:10.1007/s00044-013-0471-1

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Huang, N., Shoichet, B. K., & Irwin, J. J. (2006). Benchmarking sets for molecular docking. *Journal of Medicinal Chemistry*, *49*(23), 6789–6801. doi:10.1021/jm0608356 PMID:17154509

Huang, S.-Y., Grinter, S. Z., & Zou, X. (2010). Scoring functions and their evaluation methods for protein-ligand docking: Recent advances and future directions. *Physical Chemistry Chemical Physics*, *12*(40), 12899–12908. doi:10.1039/c0cp00151a PMID:20730182

Hughes, J. P., Rees, S., Kalindjian, S. B., & Philpott, K. L. (2011). Principles of early drug discovery. *British Journal of Pharmacology*, *162*(6), 1239–1249. doi:10.1111/j.1476-5381.2010.01127.x PMID:21091654

Jahn, A., Rosenbaum, L., Hinselmann, G., & Zell, A. (2011). 4D flexible atom-pairs: An efficient probabilistic conformation space comparison for ligand-based virtual screening. *Journal of Cheminformatics*, *3*(23), 1–17. doi:10.1186/1758-2946-3-23 PMID:21214931

Jain, A. N. (2006). Scoring functions for protein-ligand docking. *Current Protein & Peptide Science*, *7*(5), 407–420. doi:10.2174/138920306778559395 PMID:17073693

Johnson, R., Streicher, E. M., Louw, G. E., Warren, R. M., van Helden, P. D., & Victor, T. C. (2006). Drug resistance in *Mycobacterium tuberculosis. Current Issues in Molecular Biology*, *8*(2), 97–111. PMID:16878362

Jones, G., Willet, P., Glen, R. C., Leach, A. R., & Taylor, R. (1997). Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology*, *267*(3), 727–748. doi:10.1006/jmbi.1996.0897 PMID:9126849

Kamaria, P., & Kawathekar, N. (2014). Ligand-based 3D-QSAR analysis and virtual screening in exploration of new scaffolds as *Plasmodium falciparum* glutathione reductase inhibitors. *Medicinal Chemistry Research*, *23*(1), 25–33. doi:10.1007/s00044-013-0603-7

Katritzky, A. R., Karelson, M., & Lobanov, V. S. (1997). QSPR as a means of predicting and understanding chemical and physical properties in terms of structure. *Pure and Applied Chemistry*, *69*(2), 245–248. doi:10.1351/pac199769020245

Kenny, B. A., Bushfield, M., Parry-Smith, D. J., Fogarty, S., & Treherne, J. M. (1998). The application of high-throughput screening to novel lead discovery. In E. Jucker (Ed.), *Progress in drug research* (pp. 245–269). Basel, Switzerland: Birkhäuser Verlag; doi:10.1007/978-3-0348-8845-5_7

Khalaf, R. A., Sheikha, G. A., Bustanji, Y., & Taha, M. O. (2010). Discovery of new cholesteryl ester transfer protein inhibitors via ligand-based pharmacophore modeling and QSAR analysis followed by synthetic exploration. *European Journal of Medicinal Chemistry*, *45*(4), 1598–1617. doi:10.1016/j.ejmech.2009.12.070 PMID:20116902

Kitchen, D. B., Decornez, H., Furr, J. R., & Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nature Reviews. Drug Discovery*, *3*(11), 935–949. doi:10.1038/nrd1549 PMID:15520816

Klebe, G. (2006). Virtual ligand screening: Strategies, perspectives and limitations. *Drug Discovery Today*, *11*(13-14), 580–594. doi:10.1016/j.drudis.2006.05.012 PMID:16793526

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, *43*(1), 59–69. doi:10.1007/BF00337288

Kroemer, R. T. (2007). Structure-based drug design: Docking and scoring. *Current Protein & Peptide Science*, *8*(4), 312–328. doi:10.2174/138920307781369382 PMID:17696866

Krovat, E. M., Fruhwirth, K. H., & Langer, T. (2005). Pharmacophore identification, *in silico* screening, and virtual library design for inhibitors of the human factor Xa. *Journal of Chemical Information and Modeling*, *45*(1), 146–159. doi:10.1021/ci049778k PMID:15667140

Kuhn, T., Willighagen, E. L., Zielesny, A., & Steinbeck, C. (2010). CDK-Taverna: An open workflow environment for cheminformatics. *BMC Bioinformatics*, *11*(1), 159. doi:10.1186/1471-2105-11-159 PMID:20346188

Kuz'min, V. E., Artemenko, A. G., Muratov, E. N., Volineckaya, I. L., Makarov, V. A., & Riabova, O. B. et al. (2007). Quantitative structure-activity relationship studies of [(biphenoxyl)propyl]isoxazole derivatives. Inhibitors of human rhinovirus 2 replication. *Journal of Medicinal Chemistry*, *50*(17), 4205–4213. doi:10.1021/jm0704806 PMID:17665898

Lahana, R. (1999). How many leads from HTS? *Drug Discovery Today*, *4*(10), 447–448. doi:10.1016/S1359-6446(99)01393-8 PMID:10481138

Langer, T., & Hoffmann, R. D. (2001). Virtual screening: An effective tool for lead structure discovery? *Current Pharmaceutical Design*, *7*(7), 509–527. doi:10.2174/1381612013397861 PMID:11375766

Laponogov, I., Pan, X.-S., Veselkov, D. A., McAuley, K. E., Fisher, L. M., & Sanderson, M. R. (2010). Structural basis of gate-DNA breakage and resealing by type II topoisomerases. *PLoS ONE*, *5*(6), e11338. doi:10.1371/journal.pone.0011338 PMID:20596531

Laponogov, I., Sohi, M. K., Veselkov, D. A., Pan, X.-S., Sawhney, R., & Thompson, A. W. et al. (2009). Structural insight into the quinolone-DNA cleavage complex of type IIA topoisomerases. *Nature Structural & Molecular Biology*, *16*(6), 667–669. doi:10.1038/nsmb.1604 PMID:19448616

Leach, A. R., Bradshaw, J., Green, D. V. S., Hann, M. M., & Delany, J. J. (1999). Implementation of a system for reagent selection and library enumeration. *Journal of Chemical Information and Computer Sciences*, *39*(6), 1161–1172. doi:10.1021/ci9904259 PMID:10614028

Lee, K.-O., Park, H.-J., Kim, Y.-H., Seo, S.-Y., Lee, Y.-S., & Moon, S.-H. et al. (2004). CoMFA and CoMSIA 3D QSAR studies on primarane cyclooxygenase-2 (COX-2) inhibitors. *Archives of Pharmacal Research*, *27*(5), 467–470. doi:10.1007/BF02980117 PMID:15202549

Leland, B. A., Christie, B. D., Nourse, J. G., Grier, D. L., Carhart, R. E., & Maffett, T. et al. (1997). Managing the combinatorial explosion. *Journal of Chemical Information and Computer Sciences*, *37*(1), 62–70. doi:10.1021/ci960088t

Levine, C., Hiasa, H., & Marians, K. J. (1998). DNA gyrase and topoisomerase IV: Biochemical activities, physiological roles during chromosome replication, and drug sensitivities. *Biochimica et Biophysica Acta*, *1400*(1-3), 29–43. doi:10.1016/S0167-4781(98)00126-2 PMID:9748489

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Lewis, R. A. (2005). A general method for exploiting QSAR models in lead optimization. *Journal of Medicinal Chemistry*, *48*(5), 1638–1648. doi:10.1021/jm049228d PMID:15743205

Lill, M. A. (2007). Multi-dimensional QSAR in drug discovery. *Drug Discovery Today*, *12*(23-24), 1013–1017. doi:10.1016/j.drudis.2007.08.004 PMID:18061879

Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, *46*(1-3), 3–26. doi:10.1016/S0169-409X(00)00129-0 PMID:11259830

Liu, D. X., Jiang, H. L., Chen, K. X., & Ji, R. Y. (1998). A new approach to design virtual combinatorial library with genetic algorithm based on 3D grid property. *Journal of Chemical Information and Computer Sciences*, *38*(2), 233–242. doi:10.1021/ci970086o

Liu, X., Vogt, I., Haque, T., & Campillos, M. (2013). HitPick: A web server for hit identification and target prediction of chemical screening. *Bioinformatics (Oxford, England)*, *29*(15), 1910–1912. doi:10.1093/bioinformatics/btt303 PMID:23716196

Liu, Z., Diebler, R. W., Chan, H. S., & Zechiedrich, L. (2009). The why and how of DNA unlinking. *Nucleic Acids Research*, *37*(3), 661–671. doi:10.1093/nar/gkp041 PMID:19240147

Lobanov, V. S., & Agrafiotis, D. K. (2002). Scalable method for the construction and analysis of virtual combinatorial libraries. *Combinatorial Chemistry & High Throughput Screening*, *5*(2), 167–178. doi:10.2174/1386207024607392 PMID:11966425

Lobell, M., Molnar, L., & Keseru, G. M. (2003). Recent advances in the prediction of blood-brain partitioning from molecular structure. *Journal of Pharmaceutical Sciences*, *92*(2), 360–370. doi:10.1002/jps.10282 PMID:12532385

Lybrand, T. P. (1995). Ligand-protein docking and rational drug design. *Current Opinion in Structural Biology*, *5*(2), 224–228. doi:10.1016/0959-440X(95)80080-8 PMID:7648325

Manallack, D. T., Pitt, W. R., Gancia, E., Montana, J. G., Livingstone, D. J., Ford, M. G., & Whitley, D. C. (2002). Selecting screening candidates for kinase and G protein-coupled receptor targets using neural networks. *Journal of Chemical Information and Computer Sciences*, *42*(5), 1256–1262. doi:10.1021/ci020267c PMID:12377017

Marcou, G., & Rognan, D. (2007). Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. *Journal of Chemical Information and Modeling*, *47*(1), 195–207. doi:10.1021/ci600342e PMID:17238265

Martel, A. M., Leeson, P. A., & Castañer, J. (1997). BAY-12-8039. Fluoroquinolone antibacterial. *Drugs of the Future*, *22*(2), 109–113.

Matrat, S., Veziris, N., Mayer, C., Jarlier, V., Truffot-Pernot, C., & Camuset, J. et al. (2006). Functional analysis of DNA gyrase mutant enzymes carrying mutations at position 88 in the A subunit in clinical strains of *Mycobacterium tuberculosis* resistant to fluoroquinolones. *Antimicrobial Agents and Chemotherapy*, *50*(12), 4170–4173. doi:10.1128/AAC.00944-06 PMID:17015625

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Mavromoustakos, T., Durdagi, S., Koukoulitsa, C., Simcic, M., Papadopoulos, M. G., Hodoscek, M., & Golic Grdadolnik, S. (2011). Strategies in the rational drug design. *Current Medicinal Chemistry*, *18*(17), 2517–2530. doi:10.2174/092986711795933731 PMID:21568895

Maxwell, A. (1997). DNA gyrase as a drug target. *Trends in Microbiology*, *5*(3), 102–109. doi:10.1016/S0966-842X(96)10085-8 PMID:9080608

Minovski, N., Perdih, A., Novic, M., & Solmajer, T. (2013). Cluster-based molecular docking study for *in silico* identification of novel 6-fluoroquinolones as potential inhibitors against *Mycobacterium tuberculosis. Journal of Computational Chemistry*, *34*(9), 790–801. doi:10.1002/jcc.23205 PMID:23280926

Minovski, N., Perdih, A., & Solmajer, T. (2012). Combinatorially-generated library of 6-fluoroquinolone analogs as potential novel antitubercular agents: A chemometric and molecular modeling assessment. *Journal of Molecular Modeling*, *18*(5), 1735–1753. doi:10.1007/s00894-011-1179-0 PMID:21833830

Minovski, N., & Šolmajer, T. (2010). Chemometrical exploration of combinatorially generated drug-like space of 6-fluoroquinolone analogs: A QSAR study. *Acta Chimica Slovenica*, *57*(3), 529–540. PMID:24061797

Minovski, N., Vračko, M., & Šolmajer, T. (2011). Quantitative structure-activity relationship study of antitubercular fluoroquinolones. *Molecular Diversity*, *15*(2), 417–426. doi:10.1007/s11030-010-9238-5 PMID:20229318

Minovski, N., Župerl, Š., Drgan, V., & Novič, M. (2013). Assessment of applicability domain for multivariate counter-propagation artificial neural network predictive models by minimum Euclidean distance space analysis: A case study. *Analytica Chimica Acta*, *759*, 28–42. doi:10.1016/j.aca.2012.11.002 PMID:23260674

Moal, I. H., Torchala, M., Bates, P. A., & Fernández-Recio, J. (2013). The scoring of poses in protein-protein docking: Current capabilities and future directions. *BMC Bioinformatics*, *14*, 286. doi:10.1186/1471-2105-14-286 PMID:24079540

Moldover, B., Solidar, A., Montgomery, C., Miziorko, H., Murphy, J., & Wyckoff, G. J. (2012). ChemVassa: A new method for identifying small molecule hits in drug discovery. *The Open Medicinal Chemistry Journal*, *6*(1), 29–34. doi:10.2174/1874104501206010029 PMID:23525139

Morris, G. M., & Lim-Wilby, M. (2008). Molecular docking. In A. Kukol (Ed.), *Methods in molecular biology: Molecular modeling of proteins* (Vol. 443, pp. 365–382). Totowa, NJ: Humana Press; doi:10.1007/978-1-59745-177-2_19

Muegge, I. (2003). Selection criteria for drug-like molecules. *Medicinal Research Reviews*, *23*(3), 302–321. doi:10.1002/med.10041 PMID:12647312

Muegge, I., Heald, S. L., & Britelli, D. (2001). Simple selection criteria for drug-like chemical matter. *Journal of Medicinal Chemistry*, *44*(12), 1841–1846. doi:10.1021/jm015507e PMID:11384230

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Musmuca, I., Caroli, A., Mai, A., Kaushik-Basu, N., Arora, P., & Ragno, R. (2010). Combining 3-D quantitative structure-activity relationship with ligand based and structure based alignment procedures for *in Silico* screening of new hepatitis C virus NS5B polymerase inhibitors. *Journal of Chemical Information and Modeling*, *50*(4), 662–676. doi:10.1021/ci9004749 PMID:20225870

Nevin, D. K., Peters, M. B., Carta, G., Fayne, D., & Lloyd, D. G. (2012). Integrated virtual screening for the identification of novel and selective peroxisome proliferator-activated receptor (PPAR) scaffolds. *Journal of Medicinal Chemistry*, *55*(11), 4978–4989. doi:10.1021/jm300068n PMID:22582973

Nicolaou, C. A., & Brown, N. (2013). Multi-objective optimization methods in drug design. *Drug Discovery Today. Technologies*, *10*(3), e427–e435. doi:10.1016/j.ddtec.2013.02.001 PMID:24050140

Novič, M., & Zupan, J. (1995). Investigation of infrared spectra-structure correlation using Kohonen and counterpropagation neural network. *Journal of Chemical Information and Computer Sciences*, *35*(3), 454–466. doi:10.1021/ci00025a013

O'Driscoll, C. (2004). A virtual space odyssey. In *Proceedings of Horizon Symposium, Charting Chemical Space: Finding New Tools to Explore Biology* (pp. 1-4). New York, NY: Nature Publishing Group. Retrieved from http://www.nature.com/horizon/chemicalspace/background/odyssey.html

OECD. (2004). *Principles for the validation for regulatory purposes of (quantitative) structure-activity relationship models*. Paris, France: OECD.

Oprea, T. I. (2000). Property distribution of drug-related chemical databases. *Journal of Computer-Aided Molecular Design*, *14*(3), 251–264. doi:10.1023/A:1008130001697 PMID:10756480

Pantel, A., Petrella, S., Matrat, S., Brossier, F., Bastian, S., & Reitter, D. et al. (2011). DNA gyrase inhibition assays are necessary to demonstrate fluoroquinolone resistance to *gyrb* mutations in *Mycobacterium tuberculosis. Antimicrobial Agents and Chemotherapy*, *55*(10), 4524–4529. doi:10.1128/AAC.00707-11 PMID:21768507

Pantel, A., Petrella, S., Veziris, N., Brossier, F., Bastian, S., & Jarlier, V. et al. (2012). Extending the definition of the GyrB quinolone resistance-determining region in *Mycobacterium tuberculosis* DNA gyrase for assessing fluoroquiniolone resistance in *M. tuberculosis. Antimicrobial Agents and Chemotherapy*, *56*(4), 1990–1996. doi:10.1128/AAC.06272-11 PMID:22290942

Pardridge, W. M. (1995). Transport of small molecules through the blood-brain barrier: Biology and methodology. *Advanced Drug Delivery Reviews*, *15*(1-3), 5–36. doi:10.1016/0169-409X(95)00003-P

Piton, J., Petrella, S., Delarue, M., André-Leroux, G., Jarlier, V., Aubry, A., & Mayer, C. (2010). Structural insights into the quinolone resistance mechanism of *Mycobacterium tuberculosis* DNA gyrase. *PLoS ONE*, *5*(8), e12245. doi:10.1371/journal.pone.0012245 PMID:20805881

Ramesha, C. S. (2000). Comment: How many leads from HTS? *Drug Discovery Today*, *5*(2), 43–44. doi:10.1016/S1359-6446(99)01444-0 PMID:10652452

Rastelli, G., Degliesposti, G., Del Rio, A., & Sgobba, M. (2009). Binding estimation after refinement, a new automated procedure for the refinement and rescoring of docked ligands in virtual screening. *Chemical Biology & Drug Design*, *73*(3), 283–286. doi:10.1111/j.1747-0285.2009.00780.x PMID:19207463

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Rawlins, M. D. (2004). Cutting the cost of drug development? *Nature Reviews. Drug Discovery*, *3*(4), 360–364. doi:10.1038/nrd1347 PMID:15060531

Raymond, J.-L., van Deursen, R., Blum, L. C., & Ruddigkeit, L. (2010). Chemical space as a source for new drugs. *Medicinal Chemistry Communications*, *1*(1), 30–38. doi:10.1039/c0md00020e

Rishton, G. M. (1997). Reactive compounds and *in vitro* false positives in HTS. *Drug Discovery Today*, *2*(9), 382–384. doi:10.1016/S1359-6446(97)01083-0

Rishton, G. M. (2003). Non-leadlikeness and leadlikeness in biochemical screening. *Drug Discovery Today*, *8*(2), 86–96. doi:10.1016/S1359644602025722 PMID:12565011

Roy, K., Chakraborty, P., Mitra, I., Ojha, P. K., Kar, S., & Das, R. N. (2013). Some case studies on application of „r^2_m“ metrics for judging quality of quantitative structure-activity relationship predictions: emphasis on scaling of response data. *Journal of Computational Chemistry*, *34*(12), 1071–1082. doi:10.1002/jcc.23231 PMID:23299630

Sacan, A., Ekins, S., & Kortagere, S. (2012). Applications and limitations of in silico models in drug discovery. In R. S. Larson (Ed.), Bioinformatics and drug discovery, methods in molecular biology (pp. 87-124). New York, NY: Springer Science + Business Media. doi:10.1007/978-1-61779-965-5_6

Sahigara, F., Mansouri, K., Ballabio, D., Mauri, A., Consonni, V., & Todeschini, R. (2012). Comparison of different approaches to define the applicability domain of QSAR models. *Molecules (Basel, Switzerland)*, *17*(12), 4791–4810. doi:10.3390/molecules17054791 PMID:22534664

Salemme, F. R., Spurlino, J., & Bone, R. (1997). Serendipity meets precision: The integration of structure-based drug design and combinatorial chemistry for efficient drug discovery. *Structure (London, England)*, *5*(3), 319–324. doi:10.1016/S0969-2126(97)00189-5 PMID:9083110

Schlegel, B., Meier, R., Laggner, C., Schnell, D., Langer, T., & Seifert, R. et al. (2007). Generation of a homology model of the human histamine H3 receptor for ligand docking and pharmacophore-based screening. *Journal of Computer-Aided Molecular Design*, *21*(8), 437–453. doi:10.1007/s10822-007-9127-x PMID:17668276

Schwalbe, T., Kadzimirisz, D., & Jas, G. (2000). Synthesis of a library of ciprofloxacin analogues by means of sequential organic synthesis in microreactors. *QSAR & Combinatorial Science*, *24*(6), 758–768. doi:10.1002/qsar.200540005

Scior, T., Medina-Franco, J. L., Do, Q.-T., Martinez-Mayorga, K., Yunes Rojas, J. A., & Bernard, P. (2009). How to recognize and workaround pitfalls in QSAR studies: A critical review. *Current Medicinal Chemistry*, *16*(32), 4297–4313. doi:10.2174/092986709789578213 PMID:19754417

Seifert, M. H. J., & Lang, M. (2007). Essential factors for successful virtual screening. *Mini Reviews in Medicinal Chemistry*, *8*(1), 63–72. doi:10.2174/138955708783331540 PMID:18220986

Serradel, M. N., Blancafort, P., & Castañer, J. (1983). DL-8280. *Drugs of the Future*, *8*(5), 395.

Shaikh, S. A., Jain, T., Sandhu, G., Latha, N., & Jayaram, B. (2007). From drug target to leads – Sketching a physicochemical pathway for lead molecule design *In Silico*. *Current Pharmaceutical Design*, *13*(34), 3454–3470. doi:10.2174/138161207782794220 PMID:18220783

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Shi, R., Zhang, J., Li, C., Kazumi, Y., & Sugawara, I. (2006). Emergence of ofloxacin resistance in *Mycobacterium tuberculosis* clinical isolates from China as determined by *gyrA* mutation analysis using denaturing high-pressure liquid chromatography and DNA sequencing. *Journal of Clinical Microbiology*, *44*(12), 4566–4568. doi:10.1128/JCM.01916-06 PMID:17035499

Shoichet, B. K. (2004). Virtual screening of chemical libraries. *Nature*, *432*(7019), 862–865. doi:10.1038/nature03197 PMID:15602552

Sissi, C., & Palumbo, M. (2010). In front of and behind the replication fork: Bacterial type IIA topoisomerases. *Cellular and Molecular Life Sciences*, *67*(12), 2001–2024. doi:10.1007/s00018-010-0299-5 PMID:20165898

Stahura, F. L., & Bajorath, J. (2004). Virtual screening methods that complement HTS. *Combinatorial Chemistry & High Throughput Screening*, *7*(4), 259–269. doi:10.2174/1386207043328706 PMID:15200375

Sugino, A., Peebles, C. L., Kreuzer, K. N., & Cozzarelli, N. R. (1977). Mechanism of action of nalidixic acid: Purification of *Escherichia coli nalA* gene product and its relationship to DNA gyrase and a novel nicking-closing enzyme. *Proceedings of the National Academy of Sciences of the United States of America*, *74*(11), 4767–4771. doi:10.1073/pnas.74.11.4767 PMID:200930

Tan, W., Mei, H., Chao, L., Liu, T., Pan, X., Shu, M., & Yang, L. (2013). Combined QSAR and molecular docking studies on predicting P-glycoprotein inhibitors. *Journal of Computer-Aided Molecular Design*, *27*(12), 1067–1073. doi:10.1007/s10822-013-9697-8 PMID:24322389

Tian, S., Sun, H., Li, Y., Pan, P., Li, D., & Hou, T. (2013). Development and evaluation of an integrated virtual screening strategy by combining molecular docking and pharmacophore searching based on multiple protein structures. *Journal of Chemical Information and Modeling*, *53*(10), 2743–2756. doi:10.1021/ci400382r PMID:24010823

Tillotson, G. S. (1996). Quinolones: Structure-activity relationships and future predictions. *Journal of Medical Microbiology*, *44*(5), 320–324. doi:10.1099/00222615-44-5-320 PMID:8636945

Topliss, J. G. (1983). *Quantitative structure-activity relationships of drugs*. New York, NY: Academic Press.

Topliss, J. G., & Edwards, R. P. (1979). Chance factors in studies of quantitative structure-activity relationships. *Journal of Medicinal Chemistry*, *22*(10), 1238–1244. doi:10.1021/jm00196a017 PMID:513071

Tretter, E. M., Schoeffler, A. J., Weisfield, S. R., & Berger, J. M. (2010). Crystal structure of the DNA gyrase GyrA N-terminal domain from Mycobacterium tuberculosis.[PubMed]. *Proteins, Structure, Function, and Bioinformatics*, *78*(2), 492–495. doi:10.1002/prot.22600

Triballeau, N., Acher, F., Brabet, I., Pin, J.-P., & Bertrand, H.-O. (2005). Virtual screening workflow development guided by the "receiver operating characteristic" curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4. *Journal of Medicinal Chemistry*, *48*(7), 2534–2547. doi:10.1021/jm049092j PMID:15801843

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Truszkowski, A., Jayaseelan, K. V., Neumann, S., Willighagen, E. L., Zielesny, A., & Steinbeck, C. (2011). New developments on the chemoinformatics open workflow environment CDK-Taverna. *Journal of Cheminformatics*, *3*(1), 54. doi:10.1186/1758-2946-3-54 PMID:22166170

Ul-Haq, Z., Usmani, S., Shamshad, H., Mahmood, U., & Halim, S. A. (2013). A combined 3D-QSAR and docking studies for the *In-silico* prediction of HIB-protease inhibitors. *Chemistry Central Journal*, *7*(88), 1–12. doi:10.1186/1752-153X-7-88 PMID:23289739

Veber, D. F., Johnson, S. R., Cheng, H.-Y., Smith, B. R., Ward, K. W., & Kopple, K. D. (2002). Molecular properties that influence the oral bioavailability of drug candidates. *Journal of Medicinal Chemistry*, *45*(12), 2615–2623. doi:10.1021/jm020017n PMID:12036371

Vedani, A., Dobler, M., & Lill, M. A. (2005). Combining protein modeling and 6D-QSAR. Simulating the binding of structurally diverse ligands to the estrogen receptor. *Journal of Medicinal Chemistry*, *48*(11), 3700–3703. doi:10.1021/jm050185q PMID:15916421

Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W., & Taylor, R. D. (2003). Improved protein-ligand docking using GOLD. *Proteins: Structure, Function, and Bioinformatics*, *52*(4), 609–623. doi:10.1002/prot.10465 PMID:12910460

Vidal, D., Garcia-Serna, R., & Mestres, J. (2011). Ligand-based approaches to In Silico pharmacology. In J. Bajorath (Ed.), Methods in molecular biology: Chemoinformatics and computational chemical biology (vol. 672, pp. 489-502). Springer Science + Business Media, LLC. doi:10.1007/978-1-60761-839-3_19

Vilar, S., Karpiak, J., & Costanzi, S. (2009). Ligand and structure-based models for the prediction of ligand-receptor affinities and virtual screening: development and application to the β_2-adrenergic receptor. *Journal of Computational Chemistry*, *31*(4), 707–720. doi:10.1002/jcc PMID:19569204

Villoutreix, B. O., Eudes, R., & Miteva, M. A. (2009). Structure-based virtual ligand screening: Recent success stories. *Combinatorial Chemistry & High Throughput Screening*, *12*(10), 1000–1016. doi:10.2174/138620709789824682 PMID:20025565

Walters, W. P., Murcko, A. A., & Murcko, M. A. (1999). Recognizing molecules with drug-like properties. *Current Opinion in Chemical Biology*, *3*(4), 384–387. doi:10.1016/S1367-5931(99)80058-1 PMID:10419858

Wang, C., Bradley, P., & Baker, D. (2007). Protein-protein docking with backbone flexibility. *Journal of Molecular Biology*, *373*(2), 503–519. doi:10.1016/j.jmb.2007.07.050 PMID:17825317

Warr, W. A. (2012). Scientific workflow systems: Pipeline pilot and KNIME. *Journal of Computer-Aided Molecular Design*, *26*(7), 801–804. doi:10.1007/s10822-012-9577-7 PMID:22644661

Wasserman, S. A., & Cozzarelli, N. R. (1986). Biochemical topology: Applications to DNA recombination and replication. *Science*, *232*(4753), 951–960. doi:10.1126/science.3010458 PMID:3010458

Wieland, T. (1997). Combinatorics of combinatorial chemistry. *Journal of Mathematical Chemistry*, *21*(2), 141–157. doi:10.1023/A:1019166201637

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Wohlkonig, A., Chan, P. F., Fosberry, A. P., Homes, P., Huang, J., & Kranz, M. et al. (2010). Structural basis of quinolone inhibition of type IIA topoisomerases and target-mediated resistance. *Nature Structural & Molecular Biology*, *17*(9), 1152–1153. doi:10.1038/nsmb.1892 PMID:20802486

Wolber, G., & Kosara, R. (2006). Pharmacophores from macromolecular complexes with LigandScout. In T. Langer & R. D. Hoffmann (Eds.), *Pharmacophores and pharmacophore searches* (pp. 131–150). New Weinheim, Germany: WILEY-VCH; doi:10.1002/3527609164.ch6

Wolber, G., & Langer, T. (2005). LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *Journal of Chemical Information and Modeling*, *45*(1), 160–169. doi:10.1021/ci049885e PMID:15667141

Wold, S. (1991). Validation of QSAR's. *Quantitative Structure-Activity Relationship*, *10*(3), 191–193. doi:10.1002/qsar.19910100302

Wold, S., & Eriksson, L. (1995). Statistical validation of QSAR results. In Chemometrics methods in molecular design (pp. 309-318). Weinheim, Germany: Wiley-VCH.

Xiang, Z. (2006). Advances in homology protein structure modeling. *Current Protein & Peptide Science*, *7*(3), 217–227. doi:10.2174/138920306777452312 PMID:16787261

Yang, J., Roy, A., & Zhang, Y. (2013). Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics (Oxford, England)*, *29*(20), 2588–2595. doi:10.1093/bioinformatics/btt447 PMID:23975762

Yang, Y., Zhang, W., Cheng, J., Tang, Y., Peng, Y., & Li, Z. (2013). Pharmacophore, 3D-QSAR, AND Bayesian model analysis for ligands binding at the benzodiazepine site of $GABA_A$ receptors: The key roles of amino group and hydrophobic sites. *Chemical Biology & Drug Design*, *81*(5), 583–590. doi:10.1111/cbdd.12100 PMID:23279907

Young, S. S., Farmen, M. W., & Rusinko, A., III. (1996). Random versus rational: Which is better for general compound screening? *Network Science*. Retrieved from http://www.netsci.org/Science/Screening/feature09.html

Zhang, S., Golbraikh, A., Oloff, S., Kohn, H., & Tropsha, A. (2006). A novel automated lazy learning QSAR (ALL-QSAR) approach: Method Development, applications, and virtual screening of chemical databases using validated ALL-QSAR models. *Journal of Chemical Information and Modeling*, *46*(5), 1984–1995. doi:10.1021/ci060132x PMID:16995729

Zhang, Y., Yang, S., Jiao, Y., Liu, H., Yuan, H., & Lu, S. et al. (2013). An integrated virtual screening approach for VEGFR-2 inhibitors. *Journal of Chemical Information and Modeling*, *53*(12), 3163–3177. doi:10.1021/ci400429g PMID:24266594

Zheng, F., Zhan, M., Huang, X., Hammed, M. D. M. A., & Zhan, C.-G. (2014). Modeling *in vitro* inhibition of butyrylcholinesterase using molecular docking, multi-linear regression and artificial neural network approaches. *Bioorganic & Medicinal Chemistry*, *22*(1), 538–549. doi:10.1016/j.bmc.2013.10.053 PMID:24290065

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Zupan, J., & Gasteiger, J. (1999). *Neural networks in chemistry and drug design* (2nd ed.). Weinheim, Germany: Wiley-VCH.

Zupan, J., Novič, M., & Gasteiger, J. (1995). Neural networks with counter-propagation learning strategy used for modelling. *Chemometrics and Intelligent Laboratory Systems*, *27*(2), 175–187. doi:10.1016/0169-7439(95)80022-2

Zupan, J., Novič, M., & Ruisánchez, I. (1997). Kohonen and counterpropagation artificial neural networks in analytical chemistry. *Chemometrics and Intelligent Laboratory Systems*, *38*(1), 1–23. doi:10.1016/S0169-7439(97)00030-0

KEY TERMS AND DEFINITIONS

Antibacterial Agents: A common term for chemicals including synthetic or semi-synthetic drugs or other similar chemical entities that either kill or inhibit the bacterial growth.

Artificial Neural Networks: Computational non-linear modeling tools that mimic the structure and functions of the biological neural networks in the brain. During the learning process (training of the network), a complex relationship between the input and output data is established. They can be used either for numerical predictions of properties or pattern recognition purposes.

DNA Gyrase: An omnipresent molecular nanomachine from the type II DNA topoisomerase superfamily that is responsible for the unwinding of the DNA molecule during the DNA replication phase. The bacterial DNA gyrase is a well-established target of many antibacterials including nalidixic acid and their derivatives 6-fluoroquinolones.

Drug-Likeness: A qualitative measure based on a set of complex *in silico* calculable physico-chemical properties (e.g., molecular weight, logP, number of rotatable bonds, number of hydrogen bond donors and acceptors, and polar surface area) that determine whether an investigated compound is similar to the known drugs. It was found as a useful measure in the modern drug discovery and it is frequently used to filter out the so-called "drug-like" compounds from massive chemical libraries.

Protein Homology Modeling: Construction of a three-dimensional model of the "*target*" protein at atomic resolution from its amino acid sequence and an experimental three-dimensional structure of a related homologous protein ("*template*").

Quantitative Structure-Activity Relationship: An approach designed to establish relationships between the chemical structure and biological activity (or other target property) of investigated compounds in a quantitative manner.

Virtual Combinatorial Library Design: Generation of a list of structurally similar molecules in a virtual (*in silico*) environment employing the principles of combinatorial chemistry where a set of reagents (substituents or building-blocks) are specifically attached at pre-defined scaffold positions on the main structure.

Virtual Screening: A computational methodology used in the modern drug discovery to search and identify those molecular entities (small molecules) from a chemical library which are most likely to bind to a drug target, usually protein receptor or enzyme.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

ENDNOTES

1 http://www.knime.org
2 http://accelrys.com/products/pipeline-pilot
3 http://sourceforge.net/apps/mediawiki/cdk/index/.php?title=Main_Page
4 http://www.rdkit.org
5 http://ggasoftware.com/opensource
6 http://www.novamechanics.com/index.php
7 https://github.com/knime-mpicbg/HCS-Tools/wiki
8 http://www.seqan.de
9 https://www.molecular-networks.com/pipelinepilot
10 http://www.biosolveit.de/PipelinePilot
11 http://www.cresset-group.com/cresset-pipeline-pilot-component-v2-0-0-release-notes
12 http://cdktaverna.wordpress.com
13 http://www.inhibox.com
14 https://www.sciencecloud.com
15 http://accelrys.com/products/databases/sourcing/available-chemicals-directory.html
16 http://accelrys.com/products/discovery-studio/pharmacophore-ligand-based-design.html
17 http://www.inteligand.com/ligandscout
18 http://chemdb.niaid.nih.gov
19 http://www.keyorganics.ltd.uk
20 https://www.ccdc.cam.ac.uk/Solutions/GoldSuite/pages/GoldSuite.aspx
21 http://www.asinex.com
22 http://www.eyesopen.com/rocs.

This work was previously published in Quantitative Structure-Activity Relationships in Drug Design, Predictive Toxicology, and Risk Assessment edited by Kunal Roy, pages 269-317, copyright year 2015 by Medical Information Science Reference (an imprint of IGI Global).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 17
Computational Approaches for the Discovery of Novel Hepatitis C Virus NS3/4A and NS5B Inhibitors

Khac-Minh Thai
University of Medicine and Pharmacy at HCMC, Vietnam

Duy-Phong Le
University of Medicine and Pharmacy at HCMC, Vietnam

Quoc-Hiep Dong
University of Medicine and Pharmacy at HCMC, Vietnam

Minh-Tri Le
University of Medicine and Pharmacy at HCMC, Vietnam

Thi-Thanh-Lan Nguyen
University of Medicine and Pharmacy at HCMC, Vietnam

Thanh-Dao Tran
University of Medicine and Pharmacy at HCMC, Vietnam

ABSTRACT

Nonstructural 5B (NS5B) polymerase and Nonstructural 3/4A (NS3/4A) protease have proven to be promising targets for the development of anti-HCV (Hepatitis C Virus) agents. The NS5B polymerase is of paramount importance in HCV viral replication; therefore, employing NS5B inhibitors was considered an effective way for the treatment of HCV. Identifying inhibitors against NS3/4A serine protease represents another attractive approach applied in anti-HCV drug discovery, which is evidenced by its crucial role of in the biogenesis of the viral replication activity. In this chapter, many different computational approaches including Quantitative Structure-Activity Relationship (QSAR) and virtual screening in anti-HCV drug discovery were considered and discussed in detail. Virtual Screening (VS) techniques, including ligand-based and structure-based, and QSAR have been utilized for the discovery of NS5B inhibitors. Moreover, using various in silico protocols and workflows, a number of studies have been conducted with an aim of identifying potential NS3/4A blockage agents.

DOI: 10.4018/978-1-5225-0549-5.ch017

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

INTRODUCTION

Hepatitis C virus infection causes a global healthcare burden, which is likely to increase in the coming years. There are approximately 3 to 4 million new cases of HCV infection each year, and it is estimated that a minimum of 3% of worldwide population are chronically infected (Wasley et al., 2000). In most infected patients, this remarkable RNA virus causes the development of malignant chronic diseases, including cirrhosis and hepatocellular carcinoma, which often leads to liver failure and death (Alter, 1993; Hoofnagle, 1997). At present, neither an HCV vaccine nor an effective therapy against all genotypes of HCV is available. The current therapy, including pegylated interferon α, either alone or in combination with ribavirin, a broad spectrum antiviral agent, has not only limited efficacy, but also significant adverse effects (Feld et al., 2005; Fried, 2002). Hence, it is essential to develop novel agents with a high therapeutic efficacy, reduced side effects, and convenient administration to meet requirements for an anti-HCV agent.

Perceiving that urgent need, scientists around the world have conducted various research to identify as many novel HCV antiviral agents as possible. Currently, different targets for HCV therapeutic intervention encompass both structural and non-structural proteins. The structural protein is processed by host and viral proteases into four structural (core, E1, E2, and p7) and six nonstructural proteins (NS2, -3, -4A, -4B, -5A, and -5B) (Shimakami et al., 2009; Tanji et al., 1994). Non-structural protein 3/4A (serine protease - helicase) and non-structural protein 5B (RNA-dependent RNA polymerase - RdRp) have attracted the attention of medicinal chemists as targets for drug development because they play a vital role in HCV replication and the host lacks functional counterparts of them (Shimakami et al., 2009; Wang et al., 2000). This chapter will generally point out certain conducted computational approaches and virtual screening, which could be employed in combination with QSAR to build a robust model, with the goal of identifying new effective antiviral drugs utilizing NS3/4A and NS5B as principle targets.

HCV GENOME AND STRUCTURE

HCV Genome

HCV is an enveloped, positive-sense and single-stranded RNA virus approximately 9600 nucleotides in length. The significant genetic diversity was exhibited in HCV genome due to its highly error prone RNA polymerase, which makes difficulties for vaccine development and the discovery of anti–HCV agents (Francesco et al., 2005). 6 major HCV genotypes have been identified with the difference of over 30% in nucleotide sequences among each of them (Simmonds et al., 2005). RNA of HCV includes one continuous open reading frame bounded by two nontranslated regions (NTRs) at 5’ and 3’ ends. 5’ NTR constitutes the internal ribosome entry site (IRES), which plays the role as a starting point of cap-independent translation of HCV genome to produce a single polyprotein (Honda et al., 1996). The polyprotein is in turn split by both host cell and virus proteases into 10 different viral proteins with various functions and characteristics (Alvisi et al., 2011).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Structural Proteins

HCV core contains a nucleocapsid protein which fulfills numerous important functions involving RNA binding, immune modulation, signaling, and autophagy. The core protein is surrounded by HCV E1/E2 which are glycosylated envelope glycoproteins. Additionally, glycosylated proteins are able to neutralize antibody, which is attributed to ineffective immune and HCV persistence (Kaplan et al., 2007). Furthermore, there is another ion channel protein, namely p7, which is of great importance in viral assembly and release.

Non-Structural Proteins

The non-structural protein (NS) NS2 acts as the viral auto-protease, which induces the cleavage between NS2 and NS3. NS3 plays two distinct roles, as serine protease at N-terminal region and as RNA helicase at C-terminal one. NS3 protease displays a crucial function in HCV replication system by splitting downstream of NS3 at 4 site (between NS3/4a, NS4A/4B, NS4B/5A, NS5A/5B). Moreover, the TLR3 adaptor protein TRIF and mitochondrial antiviral signaling protein MAVS are also cleaved by NS3 protease, these cleavages also prevent the induction of cellular type 1 IFN pathway (Garcia-Sastre et al., 2006). In order to perform its tasks, NS3 protease requires cofactor NS4A. The function of NS4B has not been fully understood, probably associated with the formation of membranous web. The RNA viral and various host factors are bound together in a position close to HCV core by NS5A, which is a dimeric zinc-binding metalloprotein. On the other hand, NS5B is the RNA-dependent RNA polymerase (RdRp), which mediates viral genome replication (Chang et al., 2013).

COMPUTATIONAL METHOD IN DRUG DESIGN

The following part aims at introducing several concepts of computational method. QSAR is discussed in detail somewhere of this book, so, in this part, we only give a brief introduction of QSAR and mainly focus on virtual screening including pharmacophore and docking.

QSAR Modeling

Quantitative structure-activity relationship (QSAR) is a computational method which uses diverse statistical and mathematical methods to identify the correlation between bioactivity and molecular descriptors. QSAR has various applications including bioactivity, selectivity, ADMET prediction (absorption, distribution, metabolism, exertion, and toxicity). Consequently, QSAR helps medical chemists identify active or selective compounds and increase the knowledge of SAR. By carrying out QSAR study, the number of synthesized compounds would be decreased, and least possible in vitro or in vivo experiments need to be performed. Fundamentally, QSAR model building process could be divided into three steps:

1. Extraction of Molecular descriptors from structures,
2. Selection of relevant molecular descriptors, and
3. Generation of QSAR model. (Duke et al., 2006)

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

3D Pharmacophore Modeling

Paul Ehrlich was the first man to provide the definition of pharmacophore as common molecular features of ligands (Drie et al., 2007). The present concept, however, was introduced by Wemuth et al. (Wemuth et al., 1998) and was employed as IUPAC definition of pharmacophore:

A pharmacophore is the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger or block its biological response.

3D pharmacophore modeling is a reputed *in silico* technique with various advantages in drug discovery. Firstly, pharmacophore-based models provide chemical features that are supposed critical for interaction between ligand and receptor. Secondly, pharmacophore-based screening is capable of filtering out millions of compounds in reasonable time to identify a shortlist of hit compounds with various biosteric scaffolds. A significant number of books, book chapters, and reviews on pharmacophore-based virtual screening could be found in the literature, in which the most recent one is by Leach et al. (Leach et al., 2010).

Pharmacophore-based screening is a virtual screening approach that can be applied to a large number of compunds (Walter et al., 1998). It is basically used to retrieve novel lead structures from commercial or synthesizable structure databases. In addition to screening application, new biological activities or side effects are capable of being predicted by pharmacophore-based models. With the assistance of pharmacophore-based screening, high-throughput screening provides remarkably higher enrichment of active compounds, for instance, a study was conducted by Doman et al (Doman et al., 2002) which exhibited a promising result. Furthermore, integrating pharmacophore model into QSAR is another auspicious approach to generate robust models and to give elevated enrichment of true ligands.

Fundamentally, a standard pharmacophore-generating procedure includes four steps:

1. Selection of a set of ligands (only in ligand-based pharmacophore)
2. Conformational analysis,
3. Pharmacophore model generation, and
4. Pharmacophore model evaluation.

In the first step, a set of high potent ligand is selected. Then, 3D structure of ligands employed in building pharmacophore model should exhibit bioactive conformation which is able to be extracted from co-crystallographic structures. There is, however, a limited number of co-crystallographic structures available in protein data bank (PDB) website. It is abundantly evidenced that a set of diverse local minimized conformers is another effective representation of the bioactive conformers. Following that, based on the bioactive conformers of the ligands, a pharmacophore model is generated either by manual or elucidation algorithm. There are 6 kinds of model feature in pharmacophore involving hydrophobic centroids, aromatic rings, hydrogen bond acceptors or donors, cations, and anions. Manifold commercial software packages are available, such as DS CATALYST, LIGANDSCOUD, PHASE, and MOE. Finally, a new pharmacophore model is evaluated by sets of active or inactive compounds with an aim of qualifying this enrichment of active compounds. Hopefully, active compounds with diverse scaffolds will match pharmacophore model features as much as possible, and inactives will not be retrieved by the model.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

The enrichment of pharmacophore models is assessed by two basic values, namely the sensitivity (Se) and the specificity (Sp). (Triballeau et al., 2005).

$$\mathrm{Se} = \frac{\text{selected actives}}{\text{all actives}} = \frac{TP}{TP + FN}$$

$$\mathrm{Sp} = \frac{\text{discarded intives}}{\text{all inactives}} = \frac{TN}{TN + FP}$$

TP: True positive (these compounds are truly identified as positive).
TN: True negative (these compounds are truly identified as negative).
FP: False positive (these compounds are falsely identified as positive).
FN: False negative (these compounds are falsely identified as negative).

Depending on the objects, pharmacophore modeling is divided into two different approaches, namely structure-based and ligand-based pharmacophore.

Structure-Based Pharmacophore Modeling

The approach employs the information from protein-ligand interaction achieved from 3D co-crystallographic structure to generate models. Although it remains many limitations, especially protein flexibility, structure-based pharmacophore have gained more and more interest due to the growing number of protein-ligand complex crystal structures. In structure-based pharmacophore, model features usually are derived from protein-ligand interactions by using heuristics (LIGANDSCOUT) (Wolber at al., 2005), force field energy calculation in GBPM (Grid-based pharmacophore model) (Ortuso et al., 2006), or pharmacophore finger-print in FLIP (Karnachi et al., 2006).

Ligand-Based Pharmacophore

In case the co-crystallographic structure of ligand-protein complex is unavailable, known potent ligands could be used to generate pharmacophore model. In this approach, either common chemical features are identified by superimposing conformers of training set of active compounds used to build a model or a single high potent ligand is employed as a template for generating pharmacophore manually. Numerous commercial packages are available such as DS CATALYST, GALAHAD, GASP, LIGANDSCOUD, MOE, and PHASE to construct pharmacophore models. Additionally, there are two novel ligand-based pharmacophore approaches including pharmacophore fingerprint, by Young et al. (Young et al., 2006) and feature tree, by Rarey et al. (Rarey et al., 1998), which have given promising results.

Docking

Docking, a virtual screening method, includes the prediction of ligand conformations and orientation within a protein binding site and the estimation of ligand-protein interaction energy. Fundamentally, docking process involve two distinct steps, namely docking and scoring.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

In the first step, a conformational search algorithm is employed to simulate all suitable conformers of a ligand which fit into the active site of the protein. With an aim of performing conformational search task, some software packages, for instance, DOCK (Kuntz et al., 1994), CAESAR (Ehlers et al., 2007), and OMEG (Bostrom et al., 2001) generate a set of 3D possible conformers for all ligands to be docked (so-called flexibases) (Kearslry et al., 1994). In another approach, a method called incremental construction which splits a compound into many principle fragments is used. Based on geometry matching, a reference fragment is docked into the binding site. Finally, the full ligand is reconstructed incrementally by adding fragments stepwise until the whole molecule has been completed. This algorithm is implemented in many commercial software packages, such as FLEXX (Rarey et al., 1996) and SURFLEX (Jain, 2003). In the recent development of docking, two category of stochastic methods are employed including genetic algorithms and Monte Carlo (MC). In MC method, one parameter is varied at a time (torsional angle and global rotation/ translation), and the energy of generated pose is calculated. If that energy is not higher than the previous one, the conformer is kept. If energy of the conformer is higher, the probability of that pose will be identified by a selection process based on Boltzmann weight factor exp(-E/kT). MC method is usually accompanied by simulate annealing that traps the protein-ligand complex in a low energy minimum by decreasing temperature (in GLAMDOCK) (Tietze and Apostolakis., 2007). The process continuously declines the freedom of the ligand until reaching global minimum. GOLD (Jones et al., 1997) is the first docking software to use genetic algorithm (GA) in which genes in a chromosome encode conformational properties of a ligand pose. Chromosomes are evaluated by fitness function, and the strongest ones are allowed to produce offspring until a termination criterion is obtained, for example RMSD within a certain threshold. Both stochastic methods probably experience a local minimum in which global one of potential surface energy can be out of reach, however, its results usually outperform traditional methods (Douglas et al., 2004).

The second step of docking process is to calculate binding energy by scoring function. Generally, there are three kinds of scoring function including empirical, knowledge-based, and force field scoring function. In the first approach, several terms, usually polar and apolar interactions, are used to describe important properties in ligand binding in order to build a mathematic equation for predicting binding energy (Ferrara et al., 2004). Knowledge-based method employs the sum of all interatomic interactions (Helmholtz free interaction energies of protein-ligand atom pair) of protein-ligand complex to estimate the docking score. The most recent scoring function is force field-based scoring functions calculating the binding energy of protein-ligand complex on the basis of classical molecular mechanics force fields (AMBER and CHARMM). Van der Waals interaction is described by Lennard-Jones potential, wherase the electrostatic interaction is described by Coulomb energy.

Unfortunately, all three scoring functions are not very accurate in prediction of binding energy due to the omission of many factors, such as entropic component, solvation, and protein flexibility. Therefore, they are not robust enough to rank hits by elevating binding energy values. Indeed, pharmacophore modeling often outperforms docking in enrichment rate of active compound, remarkably in a very large database set (Chen et al., 2009). However, docking remains vital in drug design as a method in combination with QSAR to constructing a robust model. Docking is able to give many important features in protein-ligand binding to choose as descriptors for QSAR model or to re-evaluate high potential compounds with high predicted pIC_{50} value.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

NONSTRUCTURAL PROTEIN 5B (NS5B)

Structure of NS5B

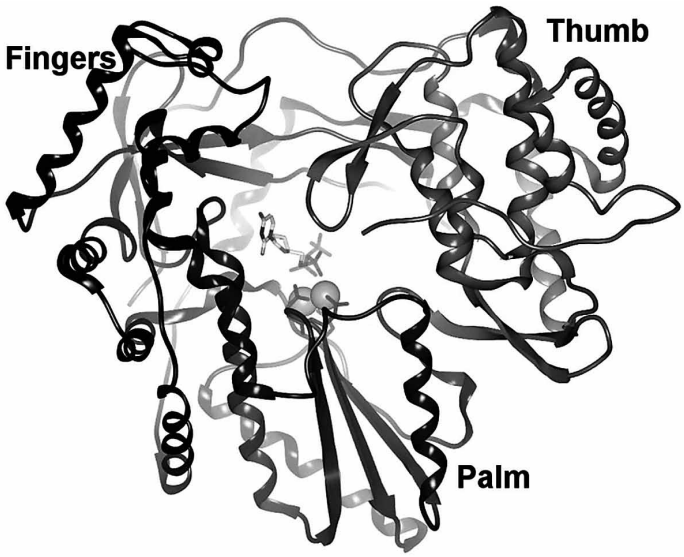
Like other RNA dependent RNA polymerases, NS5B is an error-prone enzyme, having a high error rate of its part. The viral replicase is much less accurate than the replicative prokaryotic or eucaryotic DNA polymerase due to the lack of exonuclease and proofreading domain. This also explains the reason why HCV has rapidly a wide variety of genotypes (Patel et al., 2008).

Twenty eight co-crystallographic structures of HCV NS5B have been determined to date (PDB ID: 1Z4U, 1YVF, 1YVX, 1YVZ, 2DXS, 2HWH, 2HWI, 2JCI, 2JCO, 2YOJ, 2AWZ, 2AXO, 2AXI, 2BRK, 2BRL, 2I1R, 2HA1, 2D3U, 2D3Z, 2D4I, 2GC8, 3VQ5, 3PHE, 3LKH, 3CWJ, 3CDE, 3BR9, 3OD5). It revealed that the structure of NS5B is similar to the structure of polymerase of bacteriophage Φ6 because of its fully encircled-active site (Elhefnawi et al., 2012). Like other template-dependent nucleotide polymerases, NS5B enzyme is imaginatively assimilated to the right hand. It consists of four subdomains, referred to as the palm (in violet), the thumb (in green) and the fingers (in blue) and C-terminal domain (in yellow) illustrated in (Figure 1) (T. Li et al., 2010). The palm domain contains the active site aspartates (Asp220, Asp318, Asp319) of the enzyme; whereas, the fingers and the thumb are responsible for interaction with RNA chain (Musmuca et al., 2010). The extensions of fingers and thumb lead to a more fully-enclosed active site. In the thumb domain, there is a beta-hairpin which is proposed to move upon the formation of the existing double-stranded RNA (Elhefnawi et al., 2012).

Function of NS5B

NS5B plays an essential role in HCV viral replication (Shimakami et al., 2009; Tanji et al., 1994). The key function of viral replication includes both using viral positive RNA strand as its templates and catalyzing the polymerization of ribonucleoside triphosphates (rNTP) during RNA replication (Shi-

Figure 1. Structure of NS5B (PDB ID: 2DXS)



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

makami et al., 2009). The encircled-active site of NS5B enzyme is the aspartic acid residues involved in the nucleotidyl transfer reaction during polymerization (Feld et al., 2005). Two divalent metal ions like Mg^{2+} and Mn^{2+} are coordinated with the three aspartic acid carboxylates (D220, D318 and D319), and facilitated the nucleophilic attack of 3'-hydroxy group of primer terminus or priming nucleotide on the α-phosphate of the incoming nucleotide. The ions may also be a cause of release of pyrophosphates (PPi) product (Powdrill et al., 2010).

The HCV NS5B is capable of initiating RNA synthesis de novo, especially in the absence of primer, which is proposed to be the mechanism of initiation *in vivo*. To produce the (+)-strand RNA, the positive strand RNA is used as the template to synthesize the minus strand RNA. In the initial phase, the polymerase plays a role of priming and initiating nucleotides by catalyzing the formation of a phosphodiester bond between two bound nucleotides. The next step takes place when the newly formed dinucleotides are used as a primer for addition of the third nucleotide, while the enzyme switches to the elongation mode (Powdrill et al., 2010).

The active site of NS5B controls possibly nucleotid binding and initiates the de novo RNA synthesis (Powdrill et al., 2010). Binding a structure at this active site to alter its functionality as well as intefering the function of the enzyme at each of the aforementioned individual steps by some small molecular NS5B inhibitors can inhibit viral RNA replication (Li et al., 2010; Melagraki et al., 2011).

QSAR APPROACH EMPLOYED TO IDENTIFY NOVEL NS5B INHIBITORS

2D QSAR Approach

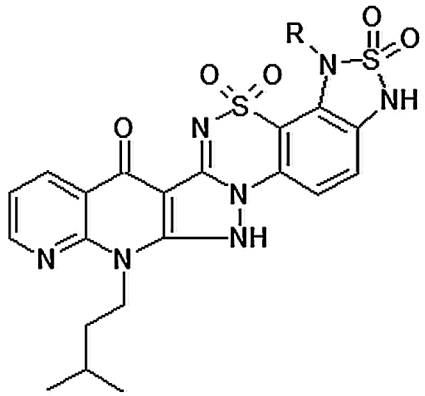
Several research into identifying NS5B inhibitors were carried out and a number of small molecular anti-NS5B compounds have been discovered. These inhibitors are categorized into two general types, namely nucleoside analogs that function as chain terminators, and non-nucleoside inhibitors (NNIs) that bind near the active site or allosteric sites on NS5B (Carroll et al., 2003; Hang et al., 2009). The NNIs include compounds belonging to benzimidazole, indole, thiophene, phenylalanine, dihydropyranone, pyranoindole, benzothiadiazine, proline sulfonamide, benzylidene, diketoacids, acrylic acid, tetracyclic indole, quinoxaline, and indole N-acetamide scaffolds (Patel et al., 2008).

QSAR approach, one of ligand-based methods, was utilized in several research to discover innovative NS5B inhibitors.

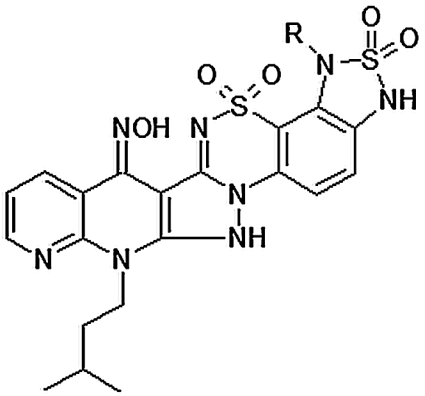
Melagraki *et al.* asserted the results of a ligand-based virtual screening on 98 compounds (benzothiadiazine analogs) which assessed as inhibitors of genotype 1 HCV polymerase (Melagraki et al., 2007). The dataset was divided into a training set of 60 compounds, and a validation set of 38 compounds based on the Kennard and Stones algorithm. Initially, quantitative structure-activity patterns were investigated for the chosen molecules. As a result, an accurate and reliable QSAR model involving five descriptors that is capable of predicting successfully the HCV inhibitory potency against genotype 1 HCV polymerase was developed. These five descriptors included lipophilicity, HOMO energy, three topological indices including Kier and Hall index order 0, 2 and 3. The model was validated employing different validation techniques. Subsequently, the structural modifications were proposed to afford novel active patterns. The effects of various structural modifications on biological activity were scrutinized and biological activities of innovative structures were estimated utilizing the constructed QSAR model. In particular, the authors described a search for optimized pharmacophore patterns by insertions, substitutions, and

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 2. Some potent NS5B inhibitors that included in the domain of applicability
Proposed from (Melagraki et al., 2007).



4-9, R=Me, Et, *n*-Pr, *i*-Pr, *n*-Bu, *s*-Bu



10-14, R=Et, *n*-Pr, Vinyl, *n*-Bu, *s*-Bu

ring fusions of pharmacophoric substituents of the main building block scaffolds. The detection of the domain of applicability defined compounds whose estimations could be accepted with confidence. Some potential NS5B inhibitors that included the domain of applicability are shown in Figure 2.

Very recently, Wang et al. developed several QSAR models for prediction of the inhibitory activity of 333 HCV NS5B polymerase inhibitors (Wang et al., 2014). All the inhibitors are NS5B NNIs fitting into the pocket of the NNI III binding site (There are totally four binding sites of NNI including Thumb domain 1, Thumb domain 2, Palm domain 1, and Palm domain 2. The first binding site is called NNI I binding site, and so forth for NNI II, NNI III, NNI IV binding site (Delang et al., 2010)). For each compound, global descriptors and 2D property autocorrelation descriptors were computed from the program ADRIANA.Code. Pearson correlation analysis was employed to choose the crucial descriptors for constructing models. The authors divided the entire dataset into a training set including 232 molecules, and a test set having 101 compounds randomly or employing a Kohonen's self-organizing map (SOM). In a next step, the randomly split training set was utilized to build two models: model 1A by multilinear regression (MLR) analysis and model 1B by support vector machine (SVM) method. The latter training set was also used to build two models: model 2A by MLR, and model 2B by SVM. The results showed that the two SVM models (model 1B and 2B) had better performance than the two MLR models (model 1A and 2A). Among these four models, model 2B performed the best. Some molecular descriptors, for instance, molecular complexity, the number of hydrogen bonding donors and the solubility of the molecule in water were found to be key factors which determined the bioactivity of the HCV NS5B inhibitors. Some other molecular properties such as electrostatic and charge properties also played significant roles in the interation between the ligand and the protein. The selected molecular descriptors were further affirmed by evaluating the interaction between two representative inhibitors and the polymerase in their crystal structures.

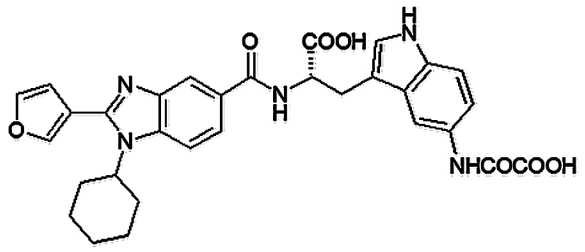
EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

3D QSAR Approach

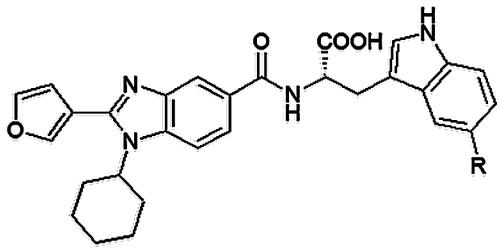
In their recent publication, Patel *et al.* conducted a study of benzimidazole derivatives as hepatitis C virus NS5B polymerase inhibitors using three-dimensional quantitative structure-activity relationship (3D-QSAR) and molecular docking method (Patel et al., 2008). They applied comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) 3D-QSAR methodologies to the training set (45 compounds) and test set (22 compounds) that include benzimidazole, tetracyclic indole, quinoxaline, and indole N-acetamide derivatives. The 3D structures of these compounds were built by the Sketch Molecule function in SYBYL. Energy minimizations were performed using the Tripos force field and Gasteiger-Marsili charges. Besides, two different alignment methods were employed. Some potent molecules used in this research are shown in Figure 3. In the ligand-based model, compound **15** (the most prospective molecule that has the lowest IC_{50} value) was selected as a template for atom based alignment in SYBYL. In the structured-based model, the molecules were aligned according to the bioactive conformations obtained from docking. In docking, all compounds were built by means of the fragment dictionary of Maestro and geometry optimized using the Optimized Potentials for Liquid Simulations-All Atom (OPLS-AA) force field. The protein was prepared and refined by SYBYL protein prepare structure tool. Then, the extra precision Glide docking method was performed to dock all molecules into the allosteric site of HCV NS5B polymerase (binding site NNI II). The final bioactive conformations were imported into a SYBYL molecular database for CoMFA and CoMSIA studies.

Figure 3. Potent NS5B inhibitors
Presented in (Patel et al., 2008).

Tryptophan derivatives

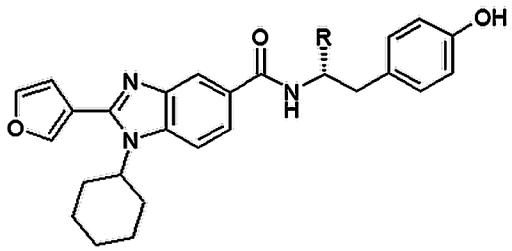


15, IC_{50} = 0.008 μM



16-19, R=OCH_2COOH, NH_2, COOH, C_5-tetrazole

Tyrosine derivatives



20-22, R=COOH, thiazol-2-amino-4-yl, N-[2-(dimethylamino)propyl]carboxamide

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

The CoMFA and CoMSIA analyses were evaluated for both the ligand- and receptor-based alignment models. However, the ligand-based alignment model caught more attention since it yielded suitably statistical data, and the conformations used in this alignment were more beneficial than those of the docked conformations. Hence, the training set for the ligand-based alignment model was examined for outliers. After removing two outlier compounds, 43 remaining compounds were reused to build 3D-QSAR models. A high bootstrapped r^2 value for CoMFA and CoMSIA models as well as a small standard deviation pointed out the existence of a similar relationship among all of the compounds in the training set. Another assessment was that the model with hydrophobic, hydrogen bond donor and acceptor fields were superior among all the model derived, which illustrated the higher predictive power of CoMSIA model than CoMFA model in the development of novel NS5B inhibitors.

Finally, the CoMFA model asserted that the tryptophan derivatives are more active than the tyrosine ones, while CoMSIA results revealed the importance of hydrogen bond donor and acceptor favorable contours at the R_1-position of the indole ring and at the R-position at the chiral carbon, respectively, for potent NS5B polymerase inhibitory activity. Docking results indicated that the N_1-cyclohexyl and the C_2-furyl substituents of the benzimidazole ring of inhibitors were binding to a deep and to a narrower hydrophobic pockets, respectively. The carbonyl oxygen atoms of the amide linkage and the −COOH group at the chiral carbon were discovered to be related to a hydrogen bond network with the guanidine of Arg503. The acidic H-bond donor group at the R_1-position of the indole ring was found to be solvent-exposed and able to forms hydrogen bonds with the backbone of Val494. The 3D-QSAR models were assessed to accurately predict the HCV NS5B polymerase inhibitory activity of structurally diverse test compounds and to provide trustworthy clues for further optimization of the benzimidazole derivatives.

Musmuca *et al.* presented the results of a combining 3D-QSAR with ligand based (LB) and structure based (SB) alignment to identify innovative molecular scaffolds as NS5B inhibitors (Musmuca et al., 2010). The study comprised 3D-QSAR model, LB and SB alignment methods, a LB-SB virtual screening and biological assessment. First, the structure based 3D-QSAR models were constructed using NS5B NNIs. A total of 24 NS5B NNIs were selected from the Protein Data Bank (PDB) as training sets, including 15 thumb and 10 palm allosteric NNIs (2JC0 complex included in both categories). All subsequent molecular docking simulations and ligand based alignments were also conducted with these training sets.

At first, the GRID/GOLPE method was employed to define two final SB 3D-QSAR models that were internally and externally validated. The internal test was cross-validation utilizing the leave-some-out with five random groups, and the external validation was using two test sets. LB and SB alignment methods were examined to evaluate the reliability on the correct molecular alignment for unknown binding mode modeled compounds. While the LB approach employed the concept of morphological similarity performed in Surflex, the SB method used Autodock program. Both Surflex and Autodock programs were capable of reproducing with minimal errors the experimental binding conformations of 24 training set compounds. 81 (thumb) and 223 (palm) modeled compounds taken from the literature were LB and SB aligned and utilized as external validation sets for 3D-QSAR models. With low error of prediction, the 3D-QSARs were proved to be useful scoring functions for the *in silico* screening procedure. The virtual screening of the NCI Diversity Set 1990 compounds led to the selection of 20 top-scoring compounds for enzymatic assays for each final model. Among the 40 chosen molecules, preliminary data provided four derivatives performing IC_{50} values ranging between 45 and 75 μM.

Structures are displayed in Table 1. Among these four compounds, NSC 123526 exhibited a docked conformation which was in good agreement with the most active compound in the thumb training set. It is a new active molecule with an unprecedented scaffold among NS5B known inhibitors.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 1. Potent HCV NS5B inhibitors

Thumb Domain	Palm Domain
36 (**NSC 123526**, IC_{50} = 46.0 μM)	**37** (**NSC 169534**, IC_{50} = 64.5 μM)
38 (**NSC 125626**, IC_{50} = 73.3 μM)	**39** (**NSC 3354**, IC_{50} = 54.3 μM)

Presented in (Musmuca et al., 2010).

VIRTUAL SCREENING METHOD IN IDENTIFYING NS5B POLYMERASE INHIBITORS

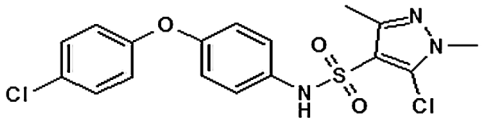
This part aims at discussing various virtual screening methods, including pharmacophore and docking, in finding anti-NS5B polymerase agents. These methods could be used either independence or combination with QSAR to create a robust model.

In their publication, Kim *et al.* presented a pharmacophore-guided virtual screening in order to identify novel anti-HCV aryl diketoacid (ADK) replacements through UNITY-based pharmacophore search of a compound library, which was then used for docking-based virtual screening (Kim et al., 2008). Based on their previous study on the binding mode analysis of the ADKs at the active site of HCV RdRp, pharmacophore models were proposed and then successfully applied to the structure-based 3D-QSAR study. LeadQuest compound library implemented in SYBYL package was employed to initiate the virtual screening procedure. Three generated pharmacophore models were utilized by UNITY search to filter the compound library. UNITY module used a conformationally flexible 3D-searching algorithm to result in rapid identification of molecules that match with the given pharmacophore. As a result, three independent compound groups with 183, 32 and 12 molecules were selected. Among 227 compounds attained from the UNITY search, only 103 molecules were successfully docked employing the FlexX-Pharm module in the SYBYL. The chosen 103 molecules with the reasonable poses were then scored utizing the Cscore™ module of SYBYL, which included five different scoring functions (Dock, Chem, FlexX, PMF, and Gold). The top 10 molecules with the highest total scores were chosen for biological assay. Among those, the compound **23** performed the most potent antiviral effect (1.26% relative cell viability in comparison with the control). The docking pose of this compound was perfectly matched with the pharmacophore model and the hydrophobic hole in the aryl-binding site. The molecule structure is shown in Figure 4.

In 2009, Ryu *et al.* discovered a new binding pocket of NS5B, distinct from the nucleotide binding site but highly conserved among diverse HCV isolates (Ryu et al., 2009). Then, they performed virtual screening of compounds that fit this binding pocket from the available database of 3.5 million molecules. As a result, it was found that the interface between thumb and palm domain was a promising binding pocket. A pharmacophore map was created using PharmoMapTM for structure-based virtual screening. The pharmacophore model illustrated the pharmacophoric features such as hydrogen donors/acceptors and hydrophobic interations, which was employed as a search query for identifying inhibitors from a 3D small molecule database. The commercially available multi-conformer 3D database includes 3.5 million compounds (PharmoDB™). After filtering out the primary screened compounds, 119 molecules were obtained for further testing *in vitro* inhibitory activity. Among those, three compounds (shown in Figure 5) with IC_{50} values of about 20 μM were identified. For a refined docking model of hit compound

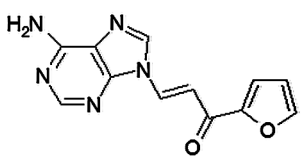
EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 4. Potent NS5B inhibitor
Proposed by (Kim et al., 2008).

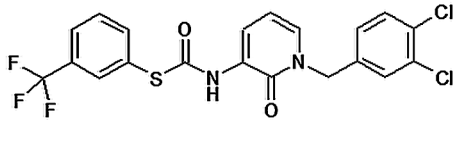


Compound **23**

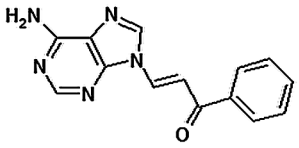
Figure 5. Chemical structure of the novel inhibitors
Proposed by (Ryu et al., 2009).



24 (CVP09011)



25 (CVP09047)



26 (CVP09081)

(compound **24**) into the HCV NS5B, automated docking simulation was carried out using AUTODOCK. X-ray crystal structure of NS5B of HCV genotype 1b from PDB (PDB ID: 2AX0) was also employed.

A new effort to discover novel polypharmacological NS5B inhibitors is described in the work of El-Hefnawi *et al.* The study used docking screening guided by contact pharmacophores and neural-network activity prediction models on 4 allosteric binding sites and molecular dynamics simulations to identify the potential hits (Elhefnawi et al., 2012). Initially, the authors employed a two-phase docking screens with Surflex and Glide Xp. The compounds were ranked in terms of scores, and important hydrogen interactions. In the next step, a machine-learning target-trained artificial neural network PIC prediction model was utilized for ranking. This provided a better correlation of IC_{50} values of the training sets for each site with distinct docking scores and sub-scores. The authors, subsequently, constructed the interaction pharmacophores-through retrospective analysis of protein-inhibitor complex X-ray structures for the interaction pharmacophore (common interation modes) of inhibitors for the four NNI binding sites. These were employed to filter the hits based on the critical binding feature of formerly reported inhibitors. This filtration process resulted in identification of prospective novel inhibitors and formerly reported ones for the thumb II and palm I sites (HCV-81) NS5B binding sites. Finally molecular dynamics simulations were conducted, confirming the binding modes and resulting in 4 leads.

The common trends of receptor-ligand interactions pharmacophores were generated using calculation from 37 PDB co-ordinates compiled according to the corresponding sites. Classifying the palm region into three sub-sites, hydrogen bonds with the backbone amide of Tyr448 and hydrophobic interaction with the deep pocket were identified as a common binding feature among all chemical subclasses of

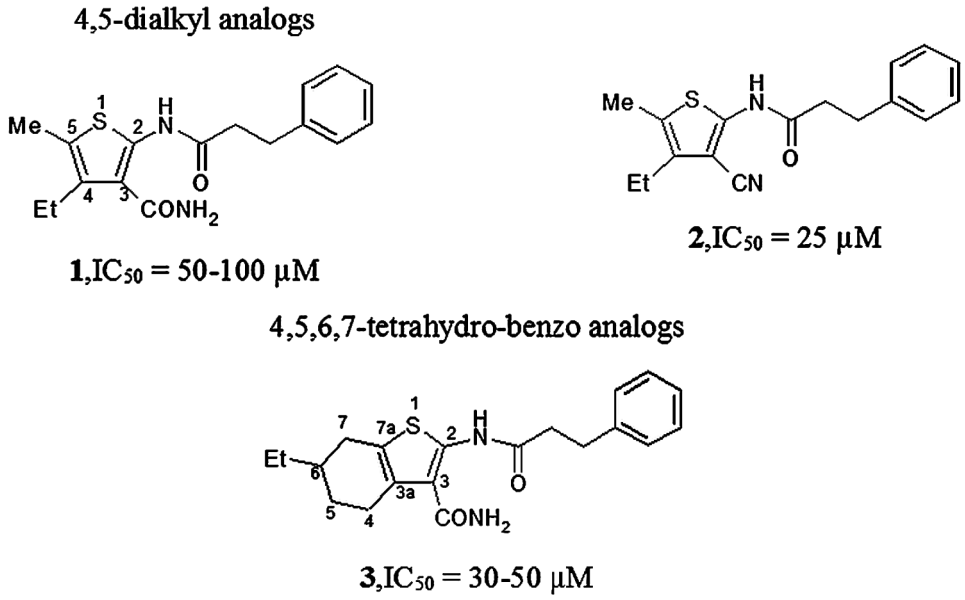
EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

palm I inhibitors. In addition, polar interaction with Asn316 was common between palm II and palm III inhibitors. While Ser476, Tyr477 and Arg501 polar contacts constituted the major features for thumb II inhibitors, Arg503 formed the only common polar interation for thumb I inhibitors. These interactions could be utilized for further study in the development of novel NS5B inhibitors.

In their publication, Louise-May *et al.* discovered novel dialkyl substituted thiophene inhibitors of HCV using *in silico* screening procedure (Louise-May et al., 2007). Available drug-like compounds were docked into the thumb domain of HCV NS5B so as to identify novel ligands. The 3D structure of HCV NS5B (1GX6) and Glide program were used for the docking. The database of these compounds was derived from an initial pool of about two million molecules from the Specs & BioSpecs, Bionet, Microsource, Available Chemicals Directory, Nanoscale, ChemDiv, Orion, Asinex, Interbio-screen, Timtec, and Chembridge databases. Tripos SLN Filters were employed to remove duplicates and reactive functionalities and to choose compounds based on the subsequent criteria: compounds having at least one ring, ≤ 1 chiral center, ≤ 8 rotatable bonds, 275-400 molecular weights, and ClogP ≤ 3.5. This procedure resulted in a library of 90,000 molecules. The top 1318 ligands with glide score of -7.17 or better were visually scrutinized to select a structural diverse set of 50 putative ligands. In the following step, the selected compounds were evaluated in an HCV replicase complex (RC) assay (cell-free) and HCV replicon assay (cell-based). The most active compounds are shown in Figure 6.

Compound **1** was followed-up by substituting at thiophene 3-, 4-, and 5-positions, and in-house synthesis of analogs. When the amide function in compound **1** was replaced by a cyano (compound **2**), there was an improvement in potency in the RC assay (IC_{50} = 25 μM). The ethyl ester and carboxylic acid at the 3-position were also tested but the results were inactive. A loss of activity was found for simple substitutions of a methyl group for an ethyl at the 4 (acyclic) or 6 (cyclic) positions, whereas activities of the 4-Et/5-Me (acyclic) and 6-Et (cyclic) substituted analogs were comparable (compounds **1** and **3**)

Figure 6. Potent HCV NS5B inhibitors
Proposed from (Louise-May et al., 2007).



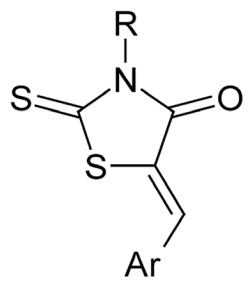
EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

in the RC assay. The active compounds in the RC assay were also assessed in the HCV replicon assay. Lower than expected EC_{50} values are partly derived from the cellular toxicity as the therapeutic window (CC_{50}/EC_{50}) for these compounds is less than 10 (3-9).

Talele *et al.*, in their recent work, identified the novel allosteric inhibitors of HCV NS5B by a combination of structure-based virtual screening, synthesis and structure-activity relationship (SAR) optimization approach (Talele et al., 2010). A ChemBridge database containing 260,000 compounds was used as the source to initiate the virtual screening procedure. Initially, the database was filtered based on Lipinski's Rule of Five, use of no tautomers, removal of metals from salts and one stereoisomer per ligand. These filtering criteria filtered out nearly 80% of the compounds from the original 2D databases to generate 52,000 3D molecules. A new database was docked in thumb II allosteric pocket (NNI II binding site) of NS5B polymerase, using the Glide's High Throughput Virtual Screening workflow. This second screening eliminated about 98% compounds and resulted in a set of remaining 650 molecules. These molecules were analyzed in context of several parameters that experience for improbable docking orientations in NNI II binding site of NS5B. Consequently, a set of 23 compounds bearing structural diversity and good docking orientations for biological assessment was yielded. Next, an *in vitro* NS5B RdRp inhibition assay was conducted, and two compounds were worth further SAR investigations. Between these two molecules, only compound **31** (shown in Table 2) had the active analogs. Following the initial rhodanine hit (compound **31**), commercially available rhodanine analogs were attained for exploration of SAR around 3- and 5-position of the rhodanine scaffold, leading to the discovery of potent analog (compound **32**). Subsequently, more than twofold enhancement of inhibitory activity was observed when 3-chlorobenzylidene was replaced with either 3,4-dichlorobenzylidene (compound **33**) or 3-phenoxybenzylidene (compound **34**), resulting in the identification of the two most potent compounds from the in-house synthetic endeavours on rhodanine analogs. Compounds **33** and **34** were then docked into each of the three HCV NS5B binding sites represented by NNI II, NNI III, and NNI IV. Further analysis of compound **34** with the tetracyclic indole- and benzylidene- binding allosteric pockets (NNI II and NNI IV, respectively) of NS5B pointed out topological similarities between these two pockets. Compound **34**, an innovative rhodanine analog with NS5B inhibitory potency in the low micromolar level range may be a promising lead for future development of more potent NS5B inhibitors.

An interesting work was indicated in the recent publication of Li *et al.* Molecular dynamics simulations, free energy decomposition and docking were employed with the aim of investigating and comparing the binding conformations of five different scaffold inhibitors (Li et al., 2010). Among many publicly available X-ray structures of HCV NS5B polymerase complexed with different NNIs, five structures

Table 2. Potent NS5B inhibitors proposed by (Talele et al., 2010)



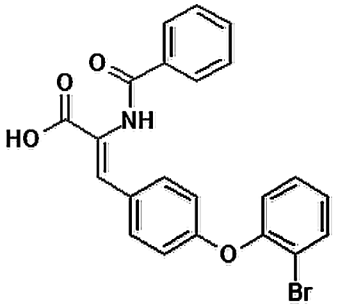
Compound	R	Ar	IC_{50} (μM)
31	$-CH_2COOH$	4-CF_3Ph	55.2 ± 1.10
32	$-CH(COOH)CH_2Ph$	2,4-ClPh	10.6 ± 1.5
33	$-CH_2COOH$	3,4-ClPh	8.4 ± 1.5
34	$-CH_2COOH$	3-PhenoxyPh	7.7 ± 1.4

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

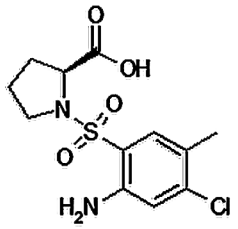
(1YVF, 2GC8, 2GIQ, 2JC0, 2QE5) were chosen based on three critical points: the protein is HCV NS5B polymerase genotype 1b, their NNIs all bind to the same palm I binding site, and these five NNIs represent different scaffolds. These NNIs included acrylic acid derivatives, proline sulfonamide derivatives, thiadiazine derivatives, acyl pyrrolidine derivatives, and anthranilic acid derivatives (Figure 7).

The binding free energies were calculated employing the Molecular Mechanics/ Generalized Born Surface Area (MM/GBSA), a molecular dynamics simulation method. The results showed that Tyr448 plays the most critical role in the binding of most inhibitors mainly through strong hydrogen bonds. Detailed analysis pointed out that major contributions which are beneficial to the binding are *van der Waals* and electrostatic energies while polar solvation energies prevent the binding. Nonpolar solvation energies contribute slightly favorably. Furthermore, several key residues for NNI II binding site were identified by the free energy decomposition with the MM/GBSA method. In the following step, the authors investigated the feasibility of docking-based drug design in this binding site. The crystal structure of protein, 2GC8, was used for further screening. After that, two diversity sets: NCI diversity set and the

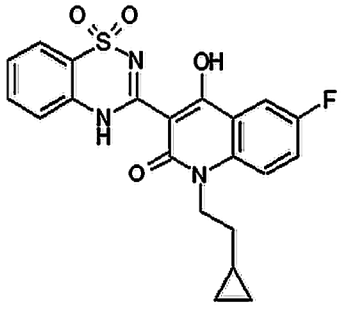
Figure 7. Structures of the five NNIs
Proposed by (T. Li et al., 2010).



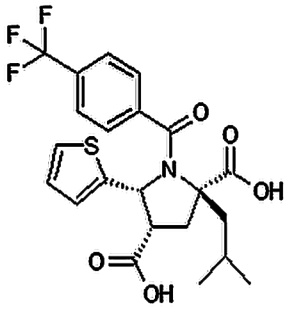
31, 1YVF (acrylic acid derivative)



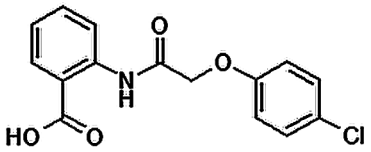
32, 2GC8 (proline sulfonamide derivative)



33, 2GIQ (thiadiazine derivative)



34, 2JC0 (acyl pyrrolidine derivative)



35, 2QE5 (anthranilic acid derivative)

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

HitFinder collection of Maybridge library were employed for virtually screening using the optimized docking protocol. The top compounds were extracted and their first-ranked binding poses were submitted to hydrogen bonding analysis. Based on the results of hydrogen bonds, molecular dynamics and free energy decomposition analysis, the authors suggested two criteria which can help to choose the candidate hits. The first one was that there are at least two stable hydrogen bonds formed between the molecule and two residues of Cys366, Ser367, Arg386, Tyr415, Gln446, Tyr448, Gln449, and Ser556. Another criterion is that tightly binding interactions are located in the deep hydrophobic pocket surrounding by Pro197, Arg200, Cys366, Met414 and Tyr448, due to a strong *van de Waals* or electrostatic interaction.

In a nut shell, there were different methods employed in drug design studies for HCV NS5B inhibitors, including construction of quantitative structure-activity relationship models, structure-based and ligand-based drug design. Although several potent NS5B inhibitors have been proposed in literature, much more efforts are needed to win the goal of rapidly identifying more effective and less toxic HCV antiviral compounds.

NS3/NS4A PROTEASE

The complete genomes of various HCV isolates were cloned and sequenced by several research groups. The structure of HCV is an enveloped, positive-strand RNA virus and its genome with a length of approximately 9.6 kilobases encodes a polypeptide precursor consisting of about 3010 amino acid residues. Cleavage of the polypeptide precursor by host cell and two viral proteases produces individual viral structural (core, envelope) and non-structural (NS) proteins. NS3 is a multifunctional enzyme with serine protease activity in the N-terminal third of the protein and an RNA helicase/NTPase activity in the C-terminal part. The serine protease domain of NS3, which requires an accessory viral protein NS4A for optimal processing, is responsible for four of five cleavage events in the non-structural region of the HCV protein. The NS3/4a protease complex has been identified as one of the most promising targets for anti-HCV therapy because it demonstrates a vital role in the replication of the HCV virus and formation of infectous viral particles.

A 2.5Å resolution structure of the serine protease domain of the NS3 with its essential cofactor NS4A indicated that the NS3/4a complex adopts a chymotrypsin/ trypsin-like fold with structurally conserved region typical of small chymotrypsin-like proteases. An eight-stranded β-barrel motif, with one of the strands made of the NS4A cofactor, is located in the N-terminal region of NS3/4a (residues 1-93 of NS3 and residues 21-34 of NS4A). The C-terminal region (residues 94-175) contains a six-stranded β-barrel with a helix as an end. The active site, which is composed of His57, Asp81 and Ser139 as the catalytic triads, located in a crevice between the two domains and formed by a shallow solvent exposed pocket requiring many interaction points for substrates or inhibitors binding. Substrates are decapeptides with 4 residues on the C terminal side and 6 residues on the N terminal residues, described as NH_2-P6-P5-P4-P3-P2-P1-P1'-P2'-P3'-P4'-OH. The pockets on the HCV NS3/4a in which the P1 site chain interacts would be called S1 pocket, and so forth for S2,S3, S4, S5, or S6 pockets and S1', S2', S3', or S4' pockets. Hence, the NS3 protease exhibits substrate specificity that requires relatively large peptides spanning the active site from S_6 and $S_{4'}$ pocket. The catalytic site contains an oxyanion hole, which stabilizes the hemiketal quaternary cleavage intermediate by hydrogen bonds with amide protons of the catalytic serine residues Ser139 and glycine Gly137. Conserved features of the substrates recognized by the NS3 protease include acidic residues in P_6 and P_5 positions, preference for cysteine in P_1 and hydrophobic residues in

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

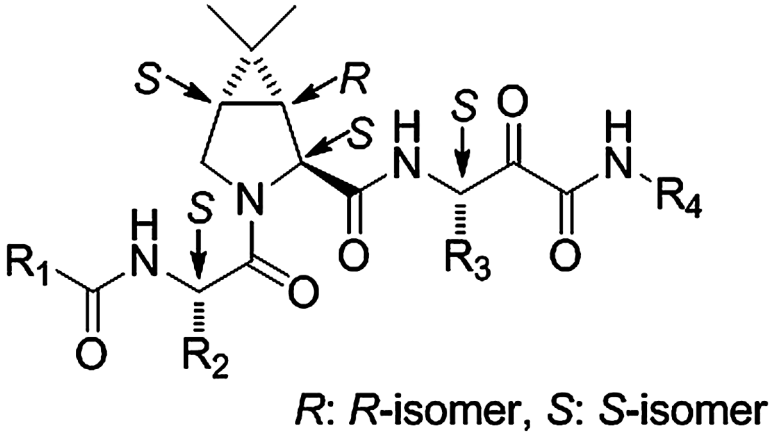
P_4. Substrates and inhibitors typically bind to the active site of the NS3 in an extended conformation and form an antiparallel β-sheet with the protease in one strand contributed by the inhibitors and the remaining strand contributed by the protease. Furthermore, a zinc ion is tetrahedrally cooridinated with Cys97, Cys 99, Cys145 and His149 residues of a site located opposite to the active site. The zinc ion plays a structural role due to the fact that the removal of the ion causes unfolding and precipitation of the protein.

CONVENTIONAL QSAR APPROACH IN DISCOVERY FOR NS3/4A PROTEASE INHIBITORS

To date, several inhibitors of NS3/NS4A with high potency against viral replication have been developed, and some of them have already undergone clinical trials, including boceprevir (SCH503034, phase III), telaprevir (VX-950, phase III), narlaprevir (SCH900518, phase III), ciluprevir (BILN 2061), danoprevir (ITMN-191), and vaniprevir (MK-7009). They all belong to peptidomimetic ketoamide group, whose general structure is shown in Figure 8, and exhibit their action by binding reversibly to the NS3 active site. Malcolm *et al.* attempted to improve the potency of these derivates by synthesizing a series of peptidomimetic ketoamide analogues. However, they have not done yet theorical structure-activity relationship (SAR) of these NS3/NS4A inhibitors and guiding researchers to design better derivatives. (Malcolm et al., 2006)

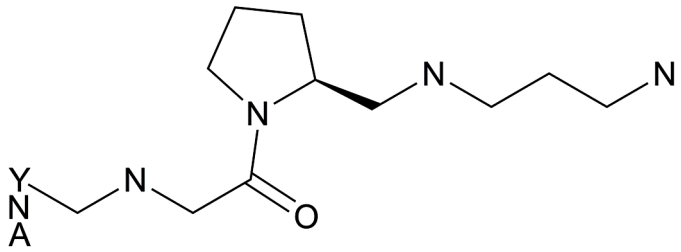
Later in 2012, Zhu *et al.* conducted quantitative structure-activity relationship (QSAR) analyses for a set of 190 narlaprevir derivatives collected from reports by Njorore's group, using comparative molecular field analysis (CoMFA), comparative molecular indices analysis (CoMSIA), hologram quantitative structure-activity relationship (HQSAR) and molecular docking simulation (Zhu et al., 2012). Of these compounds, the number of test set selected was 40 according to the values which arrayed in ascending order at regular intervals to guarantee reasonable performance. After removing a compound as an outlier due to its unreasonable low binding affinity, the final training set composed of 150 molecules. A rigid alignment was applied to superimpose all 190 compounds onto a certain common substructure, and the structure with the highest bioactivity was used as the template (Figure 9). With an aim at determining the precise binding conformations of the studied compounds, these compounds were docked into the

Figure 8. General structure of peptidomimetic ketoamide derivatives



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 9. Common substructure used for superimposing the compounds in the data set



binding cavity of NS3/NS4A. The best-docked conformations of all molecules were directly subjected to CoMFA and CoMSIA analyses.

The fields calculated in the CoMFA study included: traditional steric (Lennard-Jones, 6-12 potential) and electrostatic (Coulomb potential) fields, hydrogen bond donor and acceptor fields, and indicator steric and electrostatic fields. Among the six charge models used to generate the partial charge of the studied structures, the Gasteiger-Hückel model was identified as the best choice and used in the CoMFA and CoMSIA analyses. Partial least square (PLS) was used to correlate these fields with the biological activities, and was validated by leave one out cross validation. Eventually, the 3D-QSAR model was developed by the regression analysis with an optimum number of components obtained from the cross-validation results. The actual predictive power of the final model was then evaluated by the predictions to the external test set. As for CoMSIA study, the five following types of fields were analyzed in terms of steric, electrostatic, hydrophobic, hydrogen bond donor and acceptor. The other parameters and the calculation process were consistent with those used in the CoMFA analyses.

HQSAR emerges as a novel 2D-QSAR technique which employs specialized fragment fingerprints (molecular holograms) as predictable variables of bioactivity. Two steps including calculation of molecular fingerprints and the subsequent PLS analysis were required in building a HQSAR model.

The reported model (CoMFA1) with q^2=0.467 was not satisfactory when both of the traditional electrostatic and steric fields were taken into the PLS analysis. Upon adding the H-bond fields, however, the predictable power of the CoMFA model (q^2=0.405) got even worse. Then, the traditional steric and electrostatic fields were replaced by the indicator steric and electrostatic fields the CoMFA model was established again (CoMFA4) and CoMFA4 exhibited a better result (q^2=0.485).

In order to refine the model by enhancing the weights for those lattice points which was most pertinent to the model, the region focusing method was then used. After applying region-focusing calculations, the CoMFA model (CoMFA5) with the steric and electrostatic fields and the model (CoMFA6) with the H-bond fields showed significant improvements (q^2= 0.520 and q^2= 0.459, respectively). As for the CoMFA model (CoMFA7) with the steric, electrostatic and H-bond fields, the q^2 after applying the region-focusing calculations became much better (q^2= 0.558). More interestingly, the CoMFA model (CoMFA8) with the indicator and hydrogen bond fields was also highly promoted (q^2=0.662). As a result, the weights for the meaningful grid points could be improved and the noise could be partially eliminated.

The CoMSIA models with different combinations of fields. Compared with the CoMFA models, the CoMSIA models did not show obvious improvements (all the q^2 values were less than 0.5). The region-focusing technique was also employed to recalculate the CoMSIA fields but no improvements were observed.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Reliable CoMFA or CoMSIA models were constructed based on molecular alignment. Since the binding geometry generated by molecular docking is believed to the potential active conformation of a molecule, there are possibilities to yield more reliable structural alignment on the basis of molecular docking and then develop better CoMFA models. First, the molecular docking technique was validated by re-docking narlaprevir to the binding site of NS3/NS4a. The results showed a relatively low RMSD of 1.04 Å.

In HQSAR calculations, the effects of three parameters, including hologram length, fragment sizes, and fragment distinctions were investigated. Initially, all of the hologram lengths were collected and the default fragment size was used, and the various types of fragment distinctions were evaluated for developing robust models. It is possible that some hydrogen atoms have unusual electronic structures, thus leading to the properties that are different from those which are predefined by the HQSAR method. Therefore, in the next cycle of model optimization, the hydrogen atom type was ignored. Zhu's group also found that adding the chirality and Donor & Acceptor types, the q^2 values showed a remarkable increase. The models 1, 4 and 5 (q^2=0.542, 0.537, 0.550, respectively) were selected and optimized further by choosing different fragment sizes. After optimizing by two cycles, the best HQSAR model with q^2=0.604 was found, which employed a hologram length of 307. Then, this model was used to predict the 40 compounds in the test set. According to the prediction for the test set, both of the best CoMFA model and the best HQSAR model exhibited not very high predictive powers; for that reason, they should be amended in order to use as valuable tools in optimizing lead compound prior to chemical synthesis and biological assay.

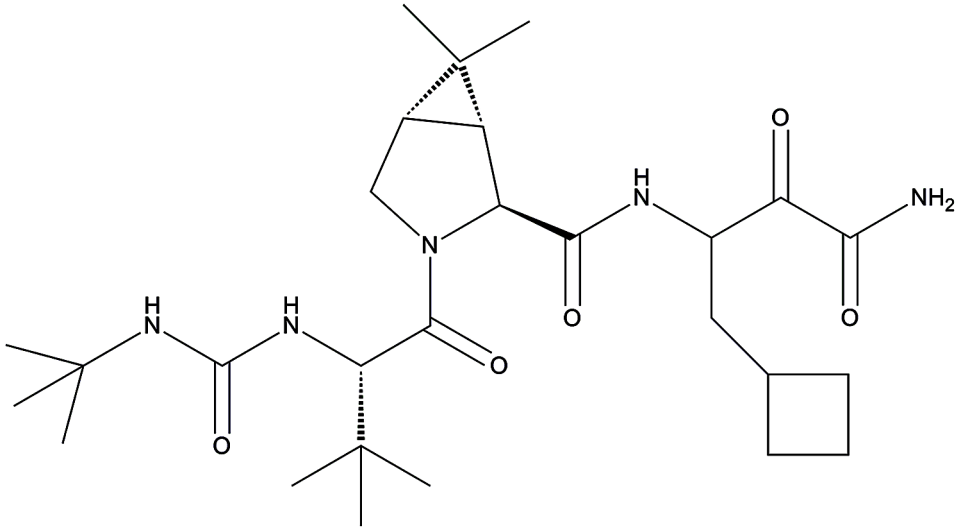
From Zhu's group results, there are two problems attached to building 3D-QSAR models of NS3/4a protease inhibitors, which are using the best docking conformation as the template and PLS as the algorithm to identify correlation between descriptors and bioactivity. Employing the best docking scoring conformation is a reasonable approach for 3D QSAR alignment template; however, it should be noticed that a major shortcoming of docking program is scoring function, which causes the best scoring function is not necessary the native one. Therefore, this solution should be used carefully based on the knowledge of X-ray complex structure, in this case is the complex of Narlaprevir and NS3/4a protease. For the latter problem, some other algorithm, such as SVM, NN, deem more reasonable method to determine the correlation between descriptors and bioactivity rather than PLS, which is only applicable in the case of linear model.

In addition to 3D-QSAR, peptidomimetic ketoamide analogues were also studied by Cunha and her group at Federal University of Lavras, using the 2D-QSAR models to discover a clearer insight into the SAR of this class of compounds.A total of 124 molecules, were based on the structure of boceprevir co-crystallized with NS3 retrieved from the Protein data Bank and their experimental inhibitory activities against HCV NS3/NS4a were prepared.

Cunha *et al.* attempted to explore the impact of substitutional variations at the P_1, $P_{1'}$, P_3 and P_4 positions shown in Figure 10 (Cunha et al., 2013). Using thermodynamic, structural and topological descriptors, including E-state decriptors, the activity data were subjected to QSAR studies. The QSAR models were developed using a training set of 93 compounds. In addition, a test set consisting of 31 structures selected from the initial 124 molecules was used for externally validating. Based on the standardized descriptor matrix, a hierarchical clustering method was applied in order to partition all compounds into ten groups. Molecules with high, medium,and low-binding affinities were arbitrary extracted from the subsets and placed in the test set. Genetic function approximation (GFA) method was used to construct QSAR models. The best model resultant of the QSAR study was constructed on the basis of these cri-

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 10. Structure of boceprevir and definition of the pharmacophoric groups (P_1, P_2, P_3 and P_4)



teria, namely, the LOO cross-validated correlation coefficient Q^2, used as the favorable parameter of model fitness; the number of significant QSAR models, and real prediction using the test set compounds (external validation). In order to identify outlier compounds, the SD of the residuals from the training set were calculated. Outliers are referred to as compounds whose residuals are more than twice the SD of the residual of fit. The results revealed that it was necessary to remove two compounds from the training set. This is another effective approach to determine outliner among traditional methods such as principle component analysis (PCA), Z-score, and self-organization map (SOM); however, it should be employed in combination with the others as mentioned above.

The squared linear correlation r^2 of 0.66 represents the best 2D-QSAR equation resulting from 93 compounds of the training set, indicating that the analyzed results exhibit a fairly good average fitness compared to the *in vitro* data. The statistical significance of the relationship between the bioactivity and the chemical structure descriptors was further demonstrated by a cross-validation analysis. LOO cross-validation (LOOcv) analysis generated a Q^2 of 0.65 with standard error of 0.36. LOO cv correlation coefficient values over 0.5 showed that the model is a useful tool for predicting affinities for new molecules in this set. The statistical significance of the relationship between the biological response and the chemical structure descriptors was further evaluated by an external validation analysis. Taking only the test set compounds into account, the predicted R^2 (R^2_{pred}) value was found to be 0.52. The standard deviation (SD) of the residual values is 0.28

As for the docking process, the most likely conformation of the ligand can be generated. The ligand conformation was identified by estimating the energy of its interaction with the target and the best (*i.e.* the lowest energy) are then returned for further analysis. The docking scoring function, the Molegro Virtual Docker (MVD), was derived from the piecewise linear potential (PLP), a simplified potential, whose parameters are fitted to protein-ligand structures and binding data scoring and further extended by the Generic Evolutionary Method.

Results from Culha et al was fairly good, but it could be strengthened by combining both QSAR and docking approaches rather than using them as separated parts. By doing this, significant interactions

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

in docking could be used as hints to select QSAR descriptors, for instance, number of hydrogen bond acceptors and donors or the surface of *van der Waals* interactions, in addition to statistical methods; furthermore, highly potent compounds predicted by QSAR model need to be re-evaluated using docking by comparing with native conformation from X-ray structure. Additionally, Culha and Zhu also proved that conventional 2D QSAR and 3D QSAR method with PLS algorithm are not successful in building a robust model for NS3/4A protease uncovalent inhibitors, SVM, genetic function approximation (GFA), or NN should be considered instead.

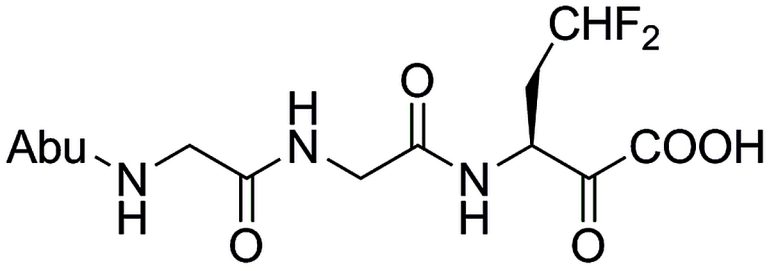
QSAR-Assisted Pharmacophore in Developing Novel Agents against NS3/4A Protease

Using amino-terminal sequence analysis of mature proteins expressed in cell culture (Wei et al., 2008), the NS3-dependent cleavage sites have been revealed. The X-ray of NS3/4a protease and determined conformation of compound 2 (BOC–GLU–LEU–FKI) (Figure 11) was employed as the template to developing structures for the residuals in the same series due to its high potency against NS3/4a protease. These molecules were divided into the training and test sets with each composing of 16 inhibitors and covering approximately the same activity range. The ligand was then docked into the active site of HCV NS3 protease using the GOLD. The covalent bond between the catalytic serine (Ser139) and the ketone group of the inhibitors was fixed during the GOLD docking process.

The electrostatic contributions at the lattice intersections where maximum steric interactions were computed were ignored. The same lattice where each molecule was submerged for CoMFA was also used for CoMSIA. In the latter, a sp^3 carbon atom of radius 1.0 Å and charge +1 was used as a probe to compute the CoMSIA similarity indices, employing the Gaussian-type distance dependence between the probe and atoms of the molecules of a data set. This functional form does not require arbitrary definition of cut-off limits and the similarity indices can be calculated at all lattice points inside and outside of a molecule.

On the basis of the docked conformation generated by the GOLD, the structure was aligned in proceeding with CoMFA and CoMSIA. The optimum number of components used to derive a model was defined as the number of components leading to the highest cross-validated r^2 (q^2) and lowest standard error of prediction. The robustness of models was judged by the conventional correlation coefficient r^2, standard error of estimate, *F*-values, and coefficient correlation between predicted and real value of external test set.

Figure 11. Compound 2 (abu: 2-aminobutyric acid)



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Still, the same training set was used for establishing some pharmacophore models by the Catalyst. All the parameters employed were set as default. There were five pharmacophore features selected for the hypothesis generation procedure, namely, HD (H-bond donor), NEG (negative charge), HY (hydrophobic), and HR (hydrophobic aromatic). The best minimized conformation of each compound in training set, adding 20 boronic acid-based inhibitors, was collected and then submitted to the hypothesis generation process.

For generating logical structures for 3D-QSAR studies, the parameters of GOLD were refined to give the minimum RMSD value between the docked and crystal structure of 1DY8. The docked structures were aligned by using several backbone atoms selected as the correspondence points. The aligned structure sets were then analyzed by the CoMFA, CoMSIA and results with significant statistics afforded are kept for further studies. The best CoMFA result generated a LOO validated q^2_{loo} of 0.749, cross-validated q^2 of 0.752 and conventional r^2 of 0.998. The CoMSIA is carried out in a stepwise manner, testing any combination of five different field indices in total, namely, steric (S), electrostatic (E), hydrophobic (H), H-bond acceptor (A) and H-bond donor (D). Values of r^2, q^2_{loo}, q^2, Standard Error of Prediction (SEP), and *F*-statistics of CoMSIA results are compared and the best result chosen in the analysis are the combination of S, E, H and D (q^2_{loo} of 0.76, the best q^2 of 0.79).

The CoMFA and CoMSIA contour maps identified both common favorable and disfavorable regions for steric interactions. The former region is located around P2 position, and the latter one covers around P3 position. The contour map generated by CoMFA agreed with CoMSIA in determining the favorable regions for negative charge interactions around position P3. However, the CoMSIA contour maps located some favorable regions around position P2, P3 and the capping group of inhibitors for lipophilic interactions. Also, the favorable region for hydrogen bond donating interactions around position P3 is determined. Based on the contour maps generated by CoMSIA, the structural features, which included HD, NEG, HY, and HR, was selected for constructing a pharmacophore model of the truncated ketoacid inhibitors. There were 10 hypotheses generated in total, then they were validated by the test set to identify the optimal model with accuracy value of 88% (Hypo1 model).

In this study, 3D QSAR model demonstrated much higher prediction ability than the model generated by Culha or Zhu in terms of better validated values (q^2_{loo}, q^2, …), which is probably due to the fact that the descriptors of compounds forming covalent bonds with NS3/4a protease has much more linear relationship with bioactivity than compounds interacts via non-covalent bonds. Besides, generating pharmacophore-based model from 3D-QSAR contour map features is considered a promising approach to enhance speed of virtual screening in large dataset of compounds, which overcome the major drawback of 3D-QSAR.

Fragment-Based Approach in Development of HCV Protease Inhibitors

Recent advances in the traditional QSAR led to introduction of fragment-based, a new branch of 2D-QSAR methodologies (Karelson et al., 2012). In this approach, the compound is treated as a system of separate units (fragments) combined in specific framework.

First step of building fragment-based model is of the calculation of the fragments', usually the FQSAR Model program is chosen. According to their definitions, these descriptors can be classified as follows:

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

1. Constitutional,
2. Geometrical,
3. Topological,
4. Charge-related,
5. Quantum chemical.

All fragments were treated by molecular mechanics, and then further optimized as random vacuum conformer with the minimum potential energy. Hence, the quantum-mechanical semi-empirical calculation in the form of the AM1 energy minimization was applied such as the Polak-Ribiere Conjugate Gradient (PRCG), Fletcher-Reeves method, or Hestenes-Stiefel approach.

Each compounds in the dataset needs to be fragmentized in certain and constant number of fragments. Further, the natural descriptors for the fragments, so-called *derived* descriptors (D_i), are used as independent variables of the main FQSAR equation in which correlation between these descriptors and bioactivity are determined by in multiple regression analysis (MLR). There are many other algorithms that can be used for constructing it instead (SVM, BPNN, GFA, Multi-layer perceptron (MLP), etc.), which are usually applied in more sophisticated cases such as non-linear relationship between descriptors and biological activity. The validation of the multi-linear model was performed using ABC validation. This method deploys three test sets in order to assess the predictability of the equation. Additionally, it takes into account the data distribution of the property.

The main power of the FQSAR algorithm is that it permits automatic construction and prediction of new compounds based on already developed FQSAR model using an internal fragments library which is automatically built by combining the fragments. The library consists of the fragments separated from the compounds in the training set or/and externally imported fragments. By posing the main criteria derived from the FQSAR equation, the fragment library can be searched for active compounds generated automatically from the assembled fragments. The assembly process includes combinatorial reconstruction of a large number of compounds using the fragments derived from the original training set. Once a new molecule is generated after each step, its properties will be predicted by the FQSAR equation.

From many reports, the studied data set of Karelson *et al.* contains 102 inhibitors of HCV NS3 protease. According to the separation, these compounds were based on several structurally different central scaffolds. All these compounds were broken into three linearly connected fragments. On the basis of the described FQSAR approach, the final fragmentation of all 102 compounds from the model development set resulted in 103 fragments. The best fragment model developed in the study had six descriptors describing $\log(1/K_i)$ in the fragment space. Validation by Randomization tests after 10,000 permutations were Y-scrambling $R^2 = 0.061$, X-scrambling $R^2 = 0.059$, XY-scrambling $R^2 = 0.062$, and cross-validation (Q^2) results were 0.799; 0.738; 0.767, respectively.

The descriptors in the FQSAR model are related to the atomic charges, electrostatic interactions and molecular weight of the compounds. The most statistically important descriptor is the maximal net atomic charge (AM1) for any atom type. This descriptor indicates that the largest charge (regardless of the atom type) in the fragment plays an essential role in the electrostatic interactions of the ligand.

The second statistically important descriptor is the polarity parameter of the molecule, which can be again ascribed to the electrostatic interactions between ligand and target. Another descriptor is the FQSAR model confirming this conjecture is the lowest coulombic interaction for N-H bonds. For these two descriptors the interactions between the fragments of the molecule is not linear and it reflects the

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

more interactive nature of these features. The feature of the ligand is probably related to the correct positioning of the inhibitor in the pockets of the receptor.

Alongside with conventional 2D-QSAR method, fragment-based drug design has emerged as a novel approach, which is proven significantly efficient in predicting the potency of NS3/4a protease inhibitors, typically in combination with high throughput screening (HTS). However, there are several problems to consider such as the size and potency of fragments, number of fragments in the library, as well as how fragments are combined. Briefly, fragments' weight should vary in the range of 150-250, a total number of hydrogen bond acceptors and donors should be below 3, cLogP approximately 3 and binding affinity ranging from nM to μM (Rees et al., 2004). Additionally, the number of fragments in the library and how they link rely on the organic synthesizing ability of a certain laboratory.

QSAR Approach in the Discovery for Multi-Target Inhibitors: The Case of HCV and HIV Co-Inhibitors

The following part aims at introducing QSAR approach in the discovery for multi-target inhibitor, including HIV-1 reverse transcriptase, HCV NS3/4a protease, and HCV NS5B polymerase.

HIV infection has become a major global public-health threat alongside HCV (Liu et al., 2011). Even more remarkable, HIV-HCV co-infection is rapidly emerging as a leading cause of morbidity and mortality all over the world, due to common rapid mutation characteristics of the two viruses as well as their similar complex influence to immunology system. Nevertheless, few researches have been carried out on the investigation of the molecular mechanism of their co-infection and designing of the multi-target co-inhibitors for HIV and HCV respectively.

In a study conducted by Liu *et al.*, a multi-target QSAR study of the inhibitors for HIV-HCV co-infection were addressed with an *in silico* machine learning technique, i.e. multi-task learning (MTL), to help to guide the co-inhibitors design. Firstly, an integrated dataset including 3 kinds of HIV target subsets and 6 kinds of HCV target subsets was employed. These molecules targeted HIV protease, integrase and reverse RT, as well as HCV NS3 and NS5B, respectively.

It should be kept in mind that the QSAR data were reported by different research groups under different protocols with different activity measurements, which make difficulty for consistency of bioactivity of inhibitors. However, these data can be integrated in an elegant multi-target QSAR relationship of the HIV-HCV co-infection, which would expect to exploit the possible synergies between different datasets and achieve a better QSAR model to guide the synthesis of certain inhibitors with improved activities for both HIV and HCV.

A novel accelerated gradient descent algorithm based MTL model was performed for multi-target QSAR modeling on the integrated datasets simultaneously. MTL has been developed in machine learning research to situations where multiple related learning tasks are accomplished together. It has been proven to be more effective than learning each task independently. The intuition underlying the framework is that the multiple related tasks can benefit each other by sharing the data and features across the tasks, which can often boost the learning performance of each single task. The QSAR modeling procedure is generally formulated as a regression model for predicting the compound activity on the basis of a given set of molecule descriptors. This study was elegantly formulated as a multi-task regression framework to reveal useful clues for multi-target drug screening and synthesizing for HIV-HCV co-infections.

Briefly, the whole process is achieved by the following steps: First, all the compounds are gathered as the input for generating their common scaffolds. If more than one scaffold is presented, they

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

are topologically aligned to produce a common numbering system. Second, assigning the scaffold by enumerating all possible substructures matches, then minimizing a pair-wise energy term which leads to the lowest possible diversity of implied *R*-group substitutes, which is expected to provide a set of analogous substitution points. Based on the pipeline, it was identified that the most efficient compound modification strategy is able to improve the molecule affinity targeting on multiple HIV-HCV enzymes. Furthermore, these identified strategies will be further explained by the joint feature ranking obtained under the multi-target QSAR paradigm.

QSAR Integrated into Structure-Based Drug Design to Discover Novel Agents against NS3/4A Protease

Recently, it has been reported that N-terminal cleavage products of substrate (typically hexapeptides) form competitive inhibitors of the NS3 protease activity (Cunha et al., 2013). These native inhibitors was employed as the template for designing novel inhibitors. Frecer and coworkers designed focused combinatorial library of hexapeptide-like inhibitors of HCV NS3 serine protease by structured-based molecular design assisted by combinatorial of the individual residues (Frecer et al., 2004). These pseudo-peptidic inhibitors were of cleavage product NS3/4A protease, which bind specifically to its active siteand their experimentally determined inhibition constants, which allowed to predict the inhibitory effect of newly designed an optimized analogues.

Crystal structures of NS3 protease with its cofactor NS4A were employed for the design of new pseudo-peptide inhibitors. The bound conformations of the designed inhibitors were modelled by adding two more residues to the initial backbone with the side-chains filling the S_6-S_1 pockets of the target's active site. In order to accommodate the side chains of the individual residue of the designed inhibitors in the pocket of the NS3/4A binding site, a complete torsion force scan of potential energy hypersurface over all the rotatable bonds of the considered side chained was conducted. All possible combinations of individual torsion angles of the side chain were generated and each structure was completely optimized using molecular mechanics. By relaxing the structures stepwise, starting with the side chains and following by protein or peptide backbone relaxation, minimizations of the enzyme-inhibitor complexes, free enzyme and free inhibitors were carried out. A sufficient number of steepest descent and conjugate gradient iterative cycles were used in all the geometry optimizations. All structures were treated as at neutral pH with the protonizable and ionizable groups being charged.

Inclusion of solvent effects in the prediction of inhibition constants improved remarkably prediction ability of enzyme-inhibitor binding, particularly for charged inhibitors. The solvation energy of the enzyme-inhibitor complexes, free enzyme and free inhibitors were computed, deploying the version of *polarized continuum model* (PCM) adapted for calculations on biopolymers.

Modeling the interactions between the enzyme and a set of 16 known inhibitors was first carried out, which is aimed at evaluating the validity of the structure-based computational process for prediction of binding affinities of newly designed structures. Due to the strongest computed interactions between compound15 and its target, it was chosen as the reference inhibitor (I_{ref}). Closer inspection of the relative interaction and contributions of solvent effect to the affinity reveals that both the enzyme-inhibitor ($\Delta\Delta H_{int}$) as well as the solute-solvent ($\Delta\Delta G_{solv}$) interaction terms are dominated by electrostatic interactions, both correlate well with the molecular charge of the inhibitor (Q_i) and $\Delta\Delta H_{int}$ and $\Delta\Delta G_{solv}$ mutually compensate their contributions to the total enzyme-inhibitor binding affinity.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

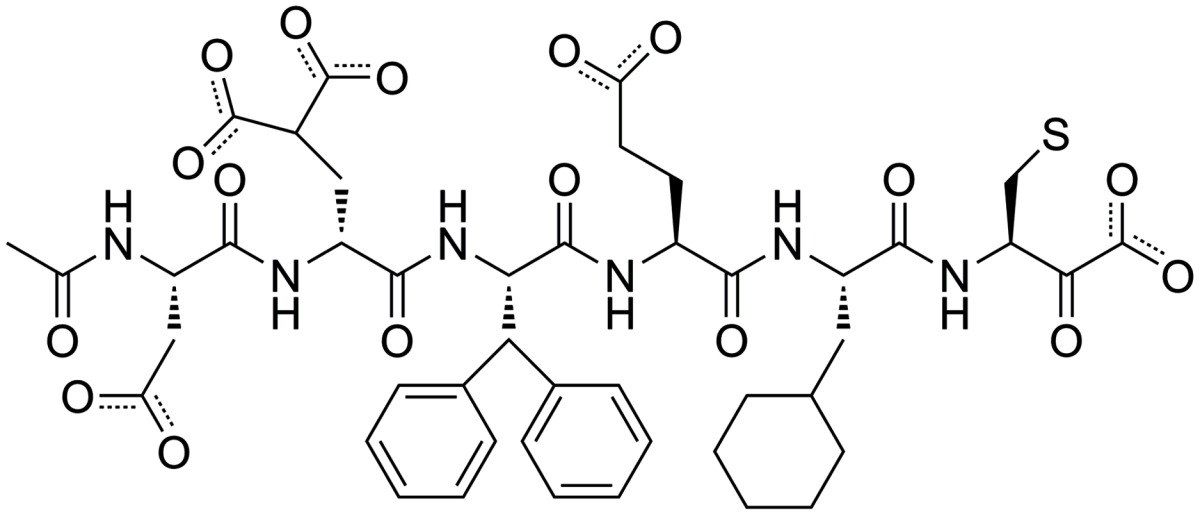
Interaction and solvent effect contributions were correlated with the experimental pK_i values using MLR algorithm. Direct correlation between the pK_i and $\Delta\Delta G_{comp}$ was not well balanced, which was caused by utilization of different computational approaches. However, a significant correlation was achieved between pK_i and the individual contributions to the relative binding affinity.

Since a good correlation between a parameter such as the molecular charge and the $\Delta\Delta G_{solv}$ contribution, the solvent term in the regression equation was replaced by the Q_i descriptor to facilitate ration structured-based design and residue optimization for a larger series (hundreds) of designed and screened peptidic inhibitors. Hence, a simple predictive QSAR model applicable to structure-based design of an inhibitor library in the simplified form was obtained:

For tight binding of peptidic inhibitors to the NS3/4A, two binding *anchors* are vital: the P_1, P_2 electrophile positioned near the catalytic site and the acidic residues in P_5, P_6 positions. Therefore, a dramatic decrease in the inhibitory potency of shorter less acidic peptides was observed after removing the acidic residues in P_5 and P_6 positions of known inhibitors. Rational design strategy of new more inhibitors should concern longer peptideswith a strong electrophile in the P_1 position. The predictive QSAR model suggests that highly acidic hexapeptides, which strongly interacts with the cationic biding site of the enzyme can form tight binding inhibitors of the NS3/4A with estimated K^*_i in the pM range. According to the model, the stabilizing effects of the negative interaction contribution to the relative binding affinity of hexapeptides can exceed the destabilizing solvent effects upon the enzyme-inhibitors formation. Using the QSAR model derived for the training set of peptidic inhibitors, it is predicted that the derivate **N**1 (Figure 12) possesses an inhibition constant K^*_i of 0.04 nM towards the target, which suggests for **N**1 higher inhibitory potency then thạt of the most effective inhibitor **I**11 (Figure 13) considered in the training set with the K_i of 0.75 nM. The computed interaction component of the binding affinity of **N**1 towards the NS3/4a exceed by almost -40 kcal mol^{-1}, which demonstrated that further structure-based modifications of the inhibitors may lead to a discovery of more potent and more specific derivatives inhibiting the NS3/4a serine protease of HCV

From the set of 50 designed inhibitor candidates with individually optimized residues in the P_6 to P_1 positions, a set of nine best residues (building blocks) useful in specific positions of hexapeptides

Figure 12. Compound N1



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 13. Compound I11

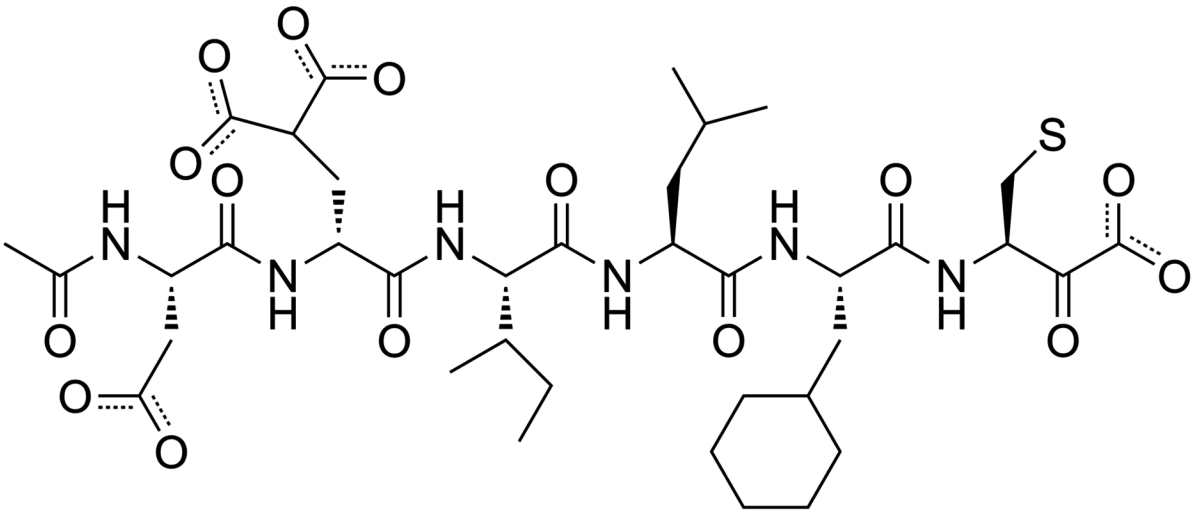
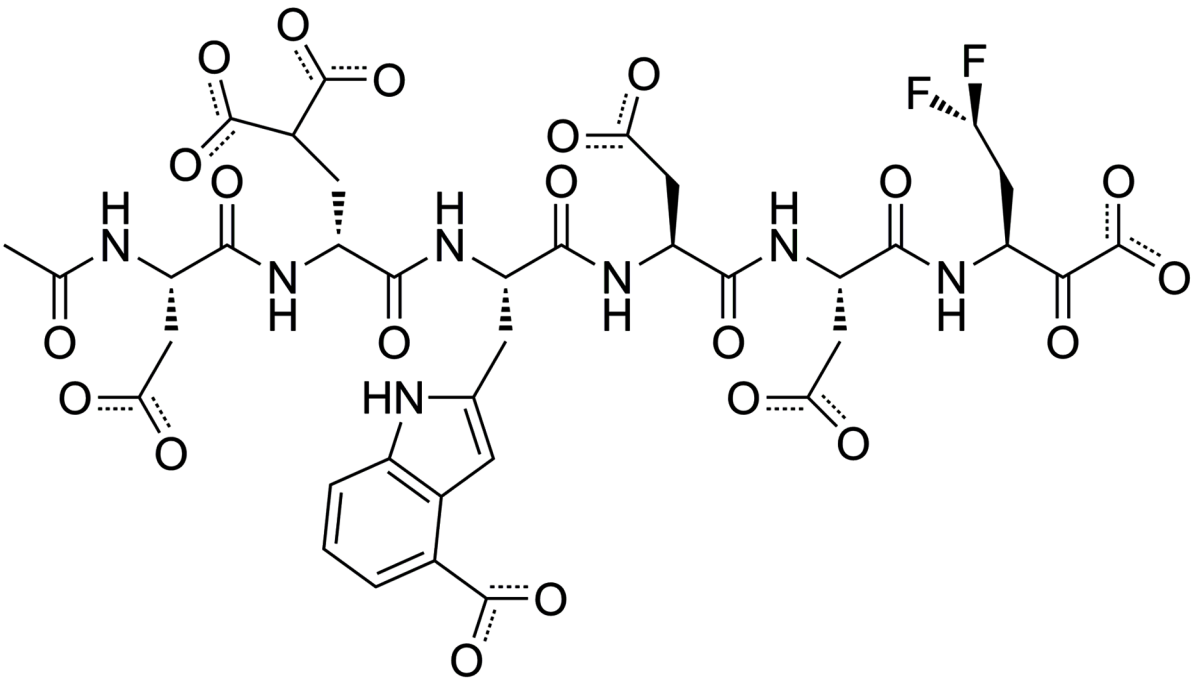


Figure 14. Compound C2



derived from **N**1 was chosen. The inhibition constants of the combinatorial analogues toward the enzyme were predicted using the QSAR model (or virtual screening), which is constructed on the basic of computed enzyme-inhibitor interaction component of the binding affinity of analogues **C**1 to **C**8 towards the NS3/4a and the estimated solvent effects that are proportional to the molecular charge Q_i of the derivatives. Compound C2 (Figure 14) shows significantly high potency against NS3/4a serine protease with the Ki of 0.3 pM.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Briefly, the innovative approach in which QSAR was integrated into structure-based model was proven significantly effective in case of designing inhibitors of NS3/4A protease. The method, however, remains several drawbacks, such as the requirement of co-crystalized structure and the approximation of calculation of ligand-target interactions. These weaknesses are capable of influencing strongly on prediction results; therefore, researchers should consider these drawbacks in prior to using the QSAR approach mentioned above.

Miscellaneous Methods in Virtual Screening NS3/4A Protease Inhibitors

Only six researches with not very good results were carried out in literature, which means QSAR alone is not robust enough in developing novel NS3/4a protease inhibitors due to the structure complexity of both inhibitor and protein. For this reason, we introduce briefly numerous successful virtual screening approaches that are capable of being combined with QSAR in order to enhance prediction results. However, this book is mainly devoted to QSAR, so we do not go with too detail into virtual screening.

Docking is a virtual screening method with increasing interest in discovering novel NS3/4A protease agents, especially compounds with diverse scaffolds rather than peptidomimetic structure such as telaprevir, danoprevir, and so on. There are two distinct methods in docking research in order to discover potent NS3/4A protease inhibitors, including docking combined with molecular dynamics or consensus docking. They all have their strength and weakness. Chaudhuri *et al.* used 5 different docking softwares, including Surflex, GOLD, Glide, Autodock, and Amber in their protocol to search for some commercial collections of compounds (Chaudhuri et al., 2012). First, Surflex was utilized to filter out compounds with inappropriate shapes. Then, GOLD was employed to dock compounds with suitable shapes, and 5000 top scored compound chosen by GOLD was redocked with three different software packs: Glide, Autodock, and Amber. A consensus list of top scoring compounds from the three softwares was generated, and 2% of 5000 compounds from the list was chosen to run *in vitro* assays. The result revealed that 12 compounds from this shortlist possess the inhibitory activity of about 50 μM with 2 compounds exhibiting remarkable promising IC_{50} values.

In another approach, only one docking algorithm was utilized in combination with molecular dynamics. In 2011, Takaya *et al.* employed a novel docking algorithm, GENIUS, involving molecular dynamics to take account into induced fit effect (Takaya et al., 2011). 0.4 million compounds, that is commercially available, were docked in order to choose 97 most promising ones to perform in vitro assay. Two compounds with inhibitory activity in micromole range, not very high active but new scaffold, were found, and their derivatives were synthesized. Another research was done by Li *et al.*, Glide was chosen as docking solfware package, and top 2000 scoring compounds were chosen to perform MM-GBSA, a kind of molecular dynamic function (Li et al., 2013). Hence, 218 interesting candidates were passed into in vitro replicon assay; then, 16 compounds were determined as NS3/4A protease inhibitor in nanomolar range.

Both of these approaches have their own advantages and drawbacks. Consensus docking-based approach above show significantly promising result, especially in searching for a new scaffold; however, it has some remarkable disadvantages. Many researches on evaluation of docking algorithms (Englebienne et al., 2009; Enyedy et al., 2008; Kim et al., 2008; Li et al., 2014; Moitessier et al., 2008; Plewczynski et al., 2011; Warren et al., 2006) revealed that no docking algorithm really outperform the others in most cases. For that reason, in some cases, only one docking algorithm is able to find active compounds from collection instead of consensus docking, so the latter could filter out compounds with true active

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

activity. Another thing is docking a large number of compounds by many docking software is really time-consuming and need a lot of labor effort. To combine with QSAR model, in the authors' point of view, the second approach is more suitable and time-reasonability. Particularly, a compound with top scoring in only one docking algorithm in consistent with being predicted high inhibitor concentration by QSAR model is reliable enough.

Pharmacophore-based virtual screening is another solution in searching for inhibitors with novel scaffolds. However, only one successful pharmacophore model has been done to date due to the complex structure of peptidomimetic compounds and protein binding pocket. Based on the crystal structure of NS3/4A protease with a macrocyclic inhibitor, manual pharmacophore model was generated including 7 key features (Wadood et al., 2014). This model was subsequently used as a filter on application to commercial databases. As a result, 786 hits was chosen for further assessment including docking and *in vitro* assay; consequently, 16 compounds were identified as NS3/4A protease inhibitors. This gave rise to some problems. The first thing is, too much pharmacophore key features were generated; another is that the research used MOEdock, which is not a highly reliable docking algorithms. Pharmacophore, however, remains a reliable method employed to strengthen QSAR model due to its ability to discard a large amount of inactive compounds.

We have already carried out a research to build a reliable model, and have obtained prominent results. Initially, we started by building a pharmacophore model that includes five key features. Then, based on 110 active compounds, a QSAR model was generated by SVM and GA algorithms. We continued our study by searching on ChemDiv databases with more than 6000 compounds matching the resultant pharmacophore model. After that, self-organization maps were built to identify which compounds are in the chemical space capable of being predicted by our QSAR model with high magnitude of reliability. Finally, we performed docking research, using LeadIT, to evaluate potential compounds predicted by QSAR that have high inhibitory activity against NS3/4A protease.

Generally, docking and pharmacophore are all suitable in combining with QSAR model with an aim of identifying a novel active compound inhibiting HCV NS3/4A protease. Especially, a novel scaffold was targeted to synthesize derivatives with simpler structure than original peptidomimetic compounds like danoprevir or simeprevir. In case of a high level of structural complexity such as NS3/4A protease, combining these virtual screening methods has significantly strengthened our QSAR approach.

CONCLUSION

In brief, in the scope of this chapter we have discussed some successful applications of QSAR in developing novel HCV antivirals. More importantly, not only two main promising targets in HCV viral cycle, NS3/4A protease and NS5B reverse transcriptase are taken into account, but also how to discover their novel inhibitor agents such as QSAR, 3D-QSAR, HQSAR, fragment-based QSAR, QSAR integrated into structure-based drug design is concerned. Although further assessment is required, they remain appealing and helpful tools in drug design and the battle against HCV in particular.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

FUTURE DIRECTION

Although several potent HCV inhibitors have been suggested, it remains crucial to make more endeavors to find more effective and less toxic chemothepapeutic agents for HCV treatment. In the battle against HCV, computational approaches have been proven powerful tools to assist drug design in the time–consuming process of identifying new active lead compounds. Both structure-based and ligand-based drug design, namely, QSAR, docking studies, library searching and *in silico* screening, are exploited to replace conventional experimental screening protocols in drug discovery. However, challenges still remain when a variety of library of compounds makes it unfeasible to control the quality of database. Moreover, reliable methods need to be employed to assure the integrity and quality of models, similarity search, data mining as well as docking in virtual screening process. These problems have to be tackled to achieve the final goal of proposing potential HCV inhibitor structures to synthesize with an aid of computational studies.

ACKNOWLEDGMENT

This work was supported by the Vietnam's National Foundation for Science and Technology Development - NAFOSTED (Grant # 106.99-2012.106 to Khac-Minh Thai and Grant # 104.01.2012.78 to Thanh-Dao Tran).

REFERENCES

Alter, M. J. (1993). The detection, transmission, and outcome of hepatitis C virus infection. *Infectious Agents and Disease*, *2*(3), 155–166. PMID:8173786

Alvisi, G., Madan, V., & Bartenschlager, R. (2011). Hepatitis C virus and host cell lipids: An intimate connection. *RNA Biology*, *8*(2), 258–269. doi:10.4161/rna.8.2.15011 PMID:21593584

Bostrom, J. (2001). Reproducing the conformations of protein-bound ligands: A critical evaluation of several popularconformational searching tools. *Journal of Computer-Aided Molecular Design*, *15*(12), 1137–1152. doi:10.1023/A:1015930826903 PMID:12160095

Carroll, S. S., Tomassini, J. E., Bosserman, M., Getty, K., Stahlhut, M. W., & Eldrup, A. B. et al. (2003). Inhibition of hepatitis C virus RNA replication by 2'-modified nucleoside analogs. *The Journal of Biological Chemistry*, *278*(14), 11979–11984. doi:10.1074/jbc.M210914200 PMID:12554735

Chang, W. K., & Kyong-Mi, C. (2013). Hepatitis C virus: Virology and life cycle. *Clinical and Molecular Hepatology*, *19*(1), 17–25. doi:10.3350/cmh.2013.19.1.17 PMID:23593605

Chaudhuri, R., Lee, H., Truong, L., Torres, J., Patel, K., & Johnson, M. E. (2012). Identification of non-macrocyclic small molecule inhibitors against the NS3/4a serine protease of hepatitis C virus through in silico screening. *Journal of Chemical Information and Modeling*, *52*(8), 2245–2256. doi:10.1021/ci300177p PMID:22697413

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chen, Z., Li, H. L., Zhang, Q. J., Bao, X. G., Yu, K., & Luo, X. et al. (2009). Pharmacophore-based virtual screening versus docking-based virtual screening: A benchmark comparison against eight targets. *Acta Pharmacologica Sinica*, *30*(12), 1694–1708. doi:10.1038/aps.2009.159 PMID:19935678

da Cunha, E. F., Matos, K. S., & Ramalho, T. C. (2013). QSAR and docking studies of HCV NS3 serine protease inhibitors. *Journal of Medicinal Chemistry*, *9*(6), 774–805. doi:10.2174/1573406411309060003 PMID:23140577

De Francesco, R., & Migliaccio, G. (2005). Challenges and successes in develop-ing new therapies for hepatitis C. *Nature*, *436*(7053), 953–960. doi:10.1038/nature04080 PMID:16107835

Delang, L., Coelmont, L., & Neyts, J. (2010, April). Antiviral therapy for hepatitis C virus: Beyond the standad of care. *Viruses*, *2*(4), 826–866. doi:10.3390/v2040826 PMID:21994657

Doman, T. N., McGovern, S. L., Witherbee, B. J., Kasten, T. P., Kurumbail, R., & Stallings, W. C. et al. (2002). Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1b. *Journal of Medicinal Chemistry*, *45*(11), 2213–2221. doi:10.1021/jm010548w PMID:12014959

Douglas, B. K., Hélèn, D., John, R. F., & Jurgen, B. (2004). Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nature Reviews. Drug Discovery*, *3*(11), 935–949. doi:10.1038/nrd1549 PMID:15520816

Dudek, A. Z., Arodz, T., & Galvez, J. (2006). Computational method in developing quantitative structure-activity relationships (QSAR): A review. *Combinatorial Chemistry & High Throughput Screening*, *9*(3), 213–228. doi:10.2174/138620706776055539 PMID:16533155

Elhefnawi, M., ElGamacy, M., & Fares, M. (2012). Multiple virtual screening approaches for finding new hepatitis C virus RNA-dependent RNA polymerase inhibitors: Structure-based screens and molecular dynamics for the pursue of new poly pharmacological inhibitors. *BMC Bioinformatics*, *13*(Suppl 17), S5. doi:10.1186/1471-2105-13-S17-S5 PMID:23282180

Englebienne, P., & Moitessier, N. (2009). Docking ligands into flexible and solvated macromolecules. 4. Are popular scoring functions accurate for this class of proteins? *Journal of Chemical Information and Modeling*, *49*(6), 1568–1580. doi:10.1021/ci8004308 PMID:19445499

Enyedy, I. J., & Egan, W. J. (2008). Can we use docking and scoring for hit-to-lead optimization? *Journal of Computer-Aided Molecular Design*, *22*(3-4), 161–168. doi:10.1007/s10822-007-9165-4 PMID:18183356

Feld, J. J., & Hoofnagle, J. H. (2005). Mechanism of action of interferon and ribavirin in treatment of hepatitis C. *Nature*, *436*(7053), 967–972. doi:10.1038/nature04082 PMID:16107837

Feng, J., Sanil, A., & Young, S. S. (2006). PharmID: Pharmacophore identification using Gibbs sampling. *Journal of Chemical Information and Modeling*, *46*(3), 1352–1359. doi:10.1021/ci050427v PMID:16711754

Ferrara, P., Gohlke, H., Price, D. J., Klebe, G., & Brooks, C. L. III. (2004). Assessing scoring functions for protein–ligand interactions. *Journal of Medicinal Chemistry*, *47*(12), 3032–3047. doi:10.1021/jm030489h PMID:15163185

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Frecer, V., Kabelac, M., De Nardi, P., Pricl, S., & Miertus, S. (2004). Structure-based design of inhibitors of NS3 serine protease of hepatitis C virus. *Journal of Molecular Graphics & Modelling*, *22*(3), 209–220. doi:10.1016/S1093-3263(03)00161-X PMID:14629979

Fried, M. W. (2002). Side effects of therapy of hepatitis C and their management. *Hepatology (Baltimore, Md.)*, *36*(5Suppl 1), S237–S244. doi:10.1053/jhep.2002.36810 PMID:12407599

Garcia-Sastre, A., & Biron, C. A. (2006). Type 1 interferons and the virus-host relationship: A lesson in detente. *Science*, *312*(5775), 879–882. doi:10.1126/science.1125676 PMID:16690858

Hang, J. Q., Yang, Y., Harris, S. F., Leveque, V., Whittington, H. J., & Rajyaguru, S. et al. (2009). Slow binding inhibition and mechanism of resistance of non-nucleoside polymerase inhibitors of hepatitis C virus. *The Journal of Biological Chemistry*, *284*(23), 15517–15529. doi:10.1074/jbc.M808889200 PMID:19246450

Honda, M., Ping, L. H., Rijnbrand, R. C., Amphlett, E., Clarke, B., Rowlands, D., & Lemon, S. M. (1996). Structural requirements for initiation of translation by internal ribosome entry within genome-length hepatitis C virus RNA. *Virology*, *222*(1), 31–42. doi:10.1006/viro.1996.0395 PMID:8806485

Hoofnagle, J. H. (1997). Hepatitis C: The clinical spectrum of disease. *Hepatology (Baltimore, Md.)*, *26*(3Suppl 1), 15s–20s. doi:10.1002/hep.510260703 PMID:9305658

Jain, A. N. (2003). Surflex: Fully automatic flexible molecular docking using a molecular similarity-based search engine. *Journal of Medicinal Chemistry*, *46*(4), 499–511. doi:10.1021/jm020406h PMID:12570372

Jones, G., Willett, P., Glen, R. C., Leach, A. R., & Taylor, R. (1997). Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology*, *267*(3), 727–748. doi:10.1006/jmbi.1996.0897 PMID:9126849

Kaplan, D. E., Sugimoto, K., Newton, K., Valiga, M. E., Ikeda, F., & Aytaman, A. et al. (2007). Discordant role of CD4 T-cell response relative to neutralizing antibody and CD8 T-cell responses in acute hepatitis C. *Gastroenterology*, *132*(2), 654–666. doi:10.1053/j.gastro.2006.11.044 PMID:17258733

Karelson, M., Dobchev, D. A., Karelson, G., Tamm, T., Tamm, K., & Nikonov, A. et al. (2012). Fragment-based development of HCV protease inhibitors for the treatment of hepatitis C. *Current Computer Aided-Drug Desig*, *8*(1), 55–61. doi:10.2174/157340912799218516 PMID:22242797

Karnachi, P., & Kulkarni, A. (2006). Application of pharmacophore fingerprints to structure-based design and data mining. In *Pharmacophores and pharmacophore searches* (pp. 193–206). Weinheim: Wiley-VCH Verlag GmbH; doi:10.1002/3527609164.ch9

Kearsley, S. K., Underwood, D. J., Sheridan, R. P., & Miller, M. D. (1994). Flexibases: A way to enhance the use of molecular docking methods. *Journal of Computer-Aided Molecular Design*, *8*(5), 565–582. doi:10.1007/BF00123666 PMID:7876901

Kim, J., Kim, K. S., Kim, D. E., & Chong, Y. (2008). Identification of novel HCV RNA-dependent RNA polymerase inhibitors using pharmacophore-guided virtual screening. *Chemical Biology & Drug Design*, *72*(6), 585–591. doi:10.1111/j.1747-0285.2008.00730.x PMID:19090925

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Kim, R., & Skolnick, J. (2008). Assessment of programs for ligand binding affinity prediction. *Journal of Computational Chemistry*, *29*(8), 1316–1331. doi:10.1002/jcc.20893 PMID:18172838

Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R., & Ferrin, T. E. (1982). A geometric approach to macromolecule–ligand interactions. *Journal of Molecular Biology*, *161*(2), 269–288. doi:10.1016/0022-2836(82)90153-X PMID:7154081

Leach, A. R., Gillet, V. J., Lewis, R. A., & Taylor, R. (2010). Three-dimensional pharmacophore methods in drug discovery. *Journal of Medicinal Chemistry*, *53*(2), 539–558. doi:10.1021/jm900817u PMID:19831387

Li, J., Ehlers, T., Sutter, J., Varma-O'brien, S., & Kirchmair, J. (2007). CAESAR: A new conformer generation algorithm based on recursive buildup and local rotational symmetry consideration. *Journal of Chemical Information and Modeling*, *47*(5), 1923–1932. doi:10.1021/ci700136x PMID:17691764

Li, J., Liu, X., Li, S., Wang, Y., Zhou, N., & Luo, C. et al. (2013). Identification of novel small molecules as inhibitors of hepatitis C virus by structure-based virtual screening. *International Journal of Molecular Sciences*, *14*(11), 22845–22856. doi:10.3390/ijms141122845 PMID:24264035

Li, T., Froeyen, M., & Herdewijn, P. (2010). Insight into ligand selectivity in HCV NS5B polymerase: Molecular dynamics simulations, free energy decomposition and docking. *Journal of Molecular Modeling*, *16*(1), 49–59. doi:10.1007/s00894-009-0519-9 PMID:19466613

Li, Y., Han, L., Liu, Z., & Wang, R. (2014). Comparative assessment of scoring functions on an updated benchmark: 2. Evaluation methods and general results. *Journal of Chemical Information and Modeling*, *54*(6), 1717–1736. doi:10.1021/ci500081m PMID:24708446

Liu, Q., Zhou, H., Liu, L., Chen, X., Zhu, R., & Cao, Z. (2011). Multi-target QSAR modelling in the analysis and design of HIV-HCV co-inhibitors: An in-silico study. *BMC Bioinformatics*, *12*(1), 294. doi:10.1186/1471-2105-12-294 PMID:21774796

Louise-May, S., Yang, W., Nie, X., Liu, D., Deshpande, M. S., & Phadke, A. S. et al. (2007). Discovery of novel dialkyl substituted thiophene inhibitors of HCV by in silico screening of the NS5B RdRp. *Bioorganic & Medicinal Chemistry Letters*, *17*(14), 3905–3909. doi:10.1016/j.bmcl.2007.04.103 PMID:17512198

Malcolm, B. A., Liu, R., Lahser, F., Agrawal, S., Belanger, B., & Butkiewicz, N. et al. (2006). SCH 503034, a mechanism-based inhibitor of hepatitis C virus NS3 protease, suppresses polyprotein maturation and enhances the antiviral activity of alpha interferon in replicon cells. *Antimicrobial Agents and Chemotherapy*, *50*(3), 1013–1020. doi:10.1128/AAC.50.3.1013-1020.2006 PMID:16495264

Melagraki, G., & Afantitis, A. (2011). Ligand and structure based virtual screening strategies for hit-finding and optimization of hepatitis C virus (HCV) inhibitors. *Current Medicinal Chemistry*, *18*(17), 2612–2619. doi:10.2174/092986711795933759 PMID:21568888

Melagraki, G., Afantitis, A., Sarimveis, H., Koutentis, P. A., Markopoulos, J., & Igglessi-Markopoulou, O. (2007). Identification of a series of novel derivatives as potent HCV inhibitors by a ligand-based virtual screening optimized procedure. *Bioorganic & Medicinal Chemistry*, *15*(23), 7237–7247. doi:10.1016/j.bmc.2007.08.036 PMID:17869118

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Moitessier, N., Englebienne, P., Lee, D., Lawandi, J., & Corbeil, C. R. (2008). Towards the development of universal, fast and highly accurate docking/scoring methods: A long way to go. *British Journal of Clinical Pharmacology*, *153*(S1Suppl 1), S7–S26. doi:10.1038/sj.bjp.0707515 PMID:18037925

Musmuca, I., Caroli, A., Mai, A., Kaushik-Basu, N., Arora, P., & Ragno, R. (2010). Combining 3-D quantitative structure-activity relationship with ligand based and structure based alignment procedures for in silico screening of new hepatitis C virus NS5B polymerase inhibitors. *Journal of Chemical Information and Modeling*, *50*(4), 662–676. doi:10.1021/ci9004749 PMID:20225870

Ortuso, F., Langer, T., & Alcaro, S. (2006). GBPM: GRID-based pharmacophore model: concept and application studies to protein–protein recognition. *Bioinformatics (Oxford, England)*, *22*(12), 1449–1455. doi:10.1093/bioinformatics/btl115 PMID:16567363

Patel, P. D., Patel, M. R., Kaushik-Basu, N., & Talele, T. T. (2008). 3D QSAR and molecular docking studies of benzimidazole derivatives as hepatitis C virus NS5B polymerase inhibitors. *Journal of Chemical Information and Modeling*, *48*(1), 42–55. doi:10.1021/ci700266z PMID:18076152

Plewczynski, D., Lazniewski, M., Augustyniak, R., & Ginalski, K. (2011). Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database. *Journal of Computational Chemistry*, *32*(4), 742–755. doi:10.1002/jcc.21643 PMID:20812323

Powdrill, M. H., Bernatchez, J. A., & Götte, M. (2010). Inhibitors of the hepatitis C virus RNA-dependent RNA polymerase NS5B. *Viruses*, *2*(10), 2169–2195. doi:10.3390/v2102169 PMID:21994615

Rarey, M., & Dixon, J. S. (1998). Feature trees: A new molecular similarity measure based on tree matching. *Journal of Computer-Aided Molecular Design*, *12*(5), 471–490. doi:10.1023/A:1008068904628 PMID:9834908

Rarey, M., Kramer, B., Lengauer, T., & Klebe, G. (1996). A fast flexible docking method using an incremental construction algorithm. *Journal of Molecular Biology*, *261*(3), 470–489. doi:10.1006/jmbi.1996.0477 PMID:8780787

Rees, D. C., Congreve, M., Murray, R. C., & Carr, R. (2004). Fragment-based lead discovery. *Nature Reviews. Drug Discovery*, *3*(8), 660–672. doi:10.1038/nrd1467 PMID:15286733

Ryu, K., Kim, N. D., Choi, S. I., Han, C. K., Yoon, J. H., & No, K. T. et al. (2009). Identification of novel inhibitors of HCV RNA-dependent RNA polymerase by pharmacophore-based virtual screening and in vitro evaluation. *Bioorganic & Medicinal Chemistry*, *17*(8), 2975–2982. doi:10.1016/j.bmc.2009.03.024 PMID:19332375

Shimakami, T., Lanford, R. E., & Lemon, S. M. (2009). Hepatitis C: Recent successes and continuing challenges in the development of improved treatment modalities. *Current Opinion in Pharmacology*, *9*(5), 537–544. doi:10.1016/j.coph.2009.08.008 PMID:19762279

Simmonds, P., Bukh, J., Combet, C., Deleage, G., Enomoto, N., & Feinstone, S. et al. (2005). Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. *Hepatology (Baltimore, Md.)*, *42*(4), 962–973. doi:10.1002/hep.20819 PMID:16149085

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Takaya, D., Yamashita, A., Kamijo, K., Gomi, J., Ito, M., & Maekawa, S. et al. (2011). A new method for induced fit docking (GENIUS) and its application to virtual screening of novel HCV NS3-4A protease inhibitors. *Bioorganic & Medicinal Chemistry*, *19*(22), 6892–6905. doi:10.1016/j.bmc.2011.09.023 PMID:21992802

Talele, T. T., Arora, P., Kulkarni, S. S., Patel, M. R., Singh, S., Chudayeu, M., & Kaushik-Basu, N. (2010). Structure-based virtual screening, synthesis and SAR of novel inhibitors of hepatitis C virus NS5B polymerase. *Bioorganic & Medicinal Chemistry*, *18*(13), 4630–4638. doi:10.1016/j.bmc.2010.05.030 PMID:20627595

Tanji, Y., Hijikata, M., Hirowatari, Y., & Shimotohno, K. (1994). Hepatitis C virus polyprotein processing: Kinetics and mutagenic analysis of serine proteinase-dependent cleavage. *Journal of Virology*, *68*(12), 8418–8422. PMID:7966638

Tietze, S., & Apostolakis, J. (2007). GlamDock: Development and validation of a new docking tool on several thousand protein–ligand complexes. *Journal of Chemical Information and Modeling*, *47*(4), 1657–1672. doi:10.1021/ci7001236 PMID:17585857

Triballeau, N., Acher, F., Brabet, I., Pin, J. P., & Bertrand, H. O. (2005). Virtual screening workflow development guided by the "receiver operating characteristic" curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4. *Journal of Medicinal Chemistry*, *48*(7), 2534–2547. doi:10.1021/jm049092j PMID:15801843

van Drie, J. H. (2007). Monty Kier and the origin of the pharmacophore concept. *Internet Electronic Journal of Molecular Design*, *6*(9), 271–279.

Wadood, A., Riaz, M., Uddin, R., & ul-Haq, Z. (2014). *In silico* identification and ealuation of lads for the simultaneous inhibition of protease and helicase activities of HCV NS3/4a protease using complex based pharmacophore mapping and virtual screening. *PLoS ONE*, *9*(2), e89109. doi:10.1371/journal.pone.0089109 PMID:24551230

Walters, W. P., Stahl, M. T., & Murcko, M. A. (1998). Virtual screening: An overview. *Drug Discovery Today*, *3*(4), 160–178. doi:10.1016/S1359-6446(97)01163-X

Wang, M., Zhong, M., Yan, A., Li, L., & Yu, C. (2014). Quantitative structure and bioactivity relationship study on HCV NS5B polymerase inhibitors. *SAR and QSAR in Environmental Research*, *25*(1), 1–15. doi:10.1080/1062936X.2013.820790 PMID:24283437

Wang, Q. M., & Heinz, B. A. (2000). Recent advances in prevention and treatment of hepatitis C virus infections. *Progress in Drug Research*, *5*, 1–32. doi:10.1007/978-3-0348-8385-6_1 PMID:11127961

Warren, G. L., Andrews, C. W., Capelli, A. M., Clarke, B., LaLonde, J., & Lambert, M. H. et al. (2006). A critical assessment of docking programs and scoring functions. *Journal of Medicinal Chemistry*, *49*(20), 5912–5931. doi:10.1021/jm050362n PMID:17004707

Wasley, A., & Alter, M. J. (2000). Epidemiology of hepatitis C: Geographic differences and temporal trends. *Seminars in Liver Disease*, *20*(1), 1–16. doi:10.1055/s-2000-9506 PMID:10895428

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Wei, H. Y., Lu, C. S., & Lin, T. H. (2008). Exploring the P2 and P3 ligand binding features for hepatitis C virus NS3 protease using some 3D QSAR techniques. *Journal of Molecular Graphics & Modelling*, *26*(7), 1131–1144. doi:10.1016/j.jmgm.2007.10.005 PMID:18024210

Wermuth, C. G., Ganellin, C. R., Lindberg, P., & Mitscher, L. A. (1998). Glossary of terms used in medicinal chemistry (IUPAC recommendations 1998). *Pure and Applied Chemistry*, *70*(5), 1129–1143. doi:10.1351/pac199870051129

Wolber, G., & Langer, T. (2005). LigandScout: 3D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *Journal of Chemical Information and Modeling*, *45*(1), 160–169. doi:10.1021/ci049885e PMID:15667141

Zhu, J., Li, Y., Yu, H., Zhang, L., Mao, X., & Hou, T. (2012). Insight into the structural requirements of narlaprevir-type inhibitors of NS3/NS4A protease based on HQSAR and molecular field analyses. *Combinational Chemistry and High Throughput Screening*, *15*(6), 439–450. doi:10.2174/138620712800563918 PMID:22263860

KEY TERMS AND DEFINITIONS

Docking: A computational method used to simulate interactions between protein and ligand and to calculate free energy if interacting process.

HCV Inhibitors: Which have diverse targets including viral attachment, NS3/4A protease, NS5A, NS5B polymerase, are significant components in Hepatitis C treatment.

Hepatitis C Virus: Positive-sense single-stranded RNA virus belonging to the family Flaviviridae. Hepatitis C virus is the cause of hepatitis C, which coulds lead to Cirrhosis or Hepatitis cancer, in humans.

NS3/4A Protease: Another protein of Hepatitis C virus, involves in cleaving process in which a polyprotein is split into many nonstructural proteins, namely NS3, NS4A, NS4B, NS5A, and NS5B protein.

NS5B Polymerase: A protein of Hepatitis C virus, plays an important role in replicating process of RNA virus.

Pharmacophore: The ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger or block its biological response.

QSAR: A computational method which uses diverse statistical and mathematical methods to identify the correlation between bioactivity and molecular descriptors.

This work was previously published in Quantitative Structure-Activity Relationships in Drug Design, Predictive Toxicology, and Risk Assessment edited by Kunal Roy, pages 318-353, copyright year 2015 by Medical Information Science Reference (an imprint of IGI Global).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 18
Multimodality Medical Image Fusion Using M-Band Wavelet and Daubechies Complex Wavelet Transform for Radiation Therapy

Satishkumar S. Chavan
Don Bosco Institute of Technology, India

Sanjay N. Talbar
S.G.G.S Institute of Engineering and Technology, India

ABSTRACT

The process of enriching the important details from various modality medical images by combining them into single image is called multimodality medical image fusion. It aids physicians in terms of better visualization, more accurate diagnosis and appropriate treatment plan for the cancer patient. The combined fused image is the result of merging of anatomical and physiological variations. It allows accurate localization of cancer tissues and more helpful for estimation of target volume for radiation. The details from both modalities (CT and MRI) are extracted in frequency domain by applying various transforms and combined them using variety of fusion rules to achieve the best quality of images. The performance and effectiveness of each transform on fusion results is evaluated subjectively as well as objectively. The fused images by algorithms in which feature extraction is achieved by M-Band Wavelet Transform and Daubechies Complex Wavelet Transform are superior over other frequency domain algorithms as per subjective and objective analysis.

DOI: 10.4018/978-1-5225-0549-5.ch018

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

INTRODUCTION

Medical imaging and its applications are playing a vital role in modern health care practices. Medical imaging has variety of acquisition techniques viz. X-rays, Ultrasound Guided imaging (USG), Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Magnetic Resonance Angiography (MRA), Positron Emission Tomography (PET), Single-Photon Emission Computed Tomography (SPECT) etc. (Depeursinge, Rodriguez, Ville, & Muller, 2014). These imaging technologies are very effectively used by physicians in disease diagnosis, treatment and post treatment analysis. Medical imaging involves acquisition of images from internal organs of the body non-invasively. These medical imaging techniques are applicable to almost every organ of human body. Araki et al. (2014) has used medical imaging for cardio-logical risk analysis using quantification of coronary calcium, whereas liver disease classification is achieved using ultrasound imaging by Suri et al. (2015). Another application of medical imaging is estimation of coronary calcium volumes using Intravascular Ultrasound (IVUS) and generating a relationship with automated carotid intima-media thickness (cIMT) for atherosclerosis disease (Araki et al., 2014; Araki et al. 2015). Intravascular ultrasound (IVUS) image registration using Rigid, Affine, B-Splines and Demons techniques are presented by Araki et al. (2015). The Intravascular Ultrasound (IVUS) bulb images with varying resolutions and imaging conditions for carotid bulb localization and bulb edge detection were experimented and stronger correlation between carotid intima-media thickness (cIMT) and coronary Synergy between percutaneous coronary intervention with TAXUS and cardiac surgery (SYNTAX) score is estimated (Ikeda et al., 2014; Ikeda et al., 2015).

Medical image processing is most advanced technology which is used in almost all stages of cancer treatment starting from diagnosis till adaptive post treatment. It is helpful in locating infected tissues, in deciding treatment plan and taking corrective action based on post treatment images. It is not possible for a device (single modality) to capture all relevant and essential information in a single slice/frame. Therefore, various modalities are used to capture details. Some image acquisition machines are also manufactured to capture more than one modality images using a single machine; but these machines are very expensive and are seldom available or facility may be present at very few medical treatment centers. Obviously, most of the hospitals and medical health centers have single modality machines. Due to use of variety of single medical image modalities, one can achieve different type of information in each modality image. The information in each image can be similar and/or complementary. Each modality has its own pros and cons. For example X-rays / Computed Tomography (CT) captures details of bone structure, hard tissues, implants etc. but unable to provide information about physiological variations whereas Magnetic Resonance Imaging (MRI) images can provide pathological soft tissues and fluid texture information but cannot support bones structures (Kessler, 2006). Thus, single modality may not provide enough clinical information. However multimodality images definitely carry enough complementary information. Therefore, combining this complementary information from various modality images into single image may be added advantage or sometime almost necessary in patient management system for cancer treatment. The process may also help doctors to diagnose and treat accurately and precisely. The process of registering and combining various modality images into a better quality single image with all relevant information from different modalities is called multimodality medical image fusion. The resultant i.e. fused image may be more suitable for visual interpretation in cancer treatment provided the process should be free from noise, addition of unwanted structures and visual artifacts.

Radiation Therapy is a most important and specialized way of treating cancer patients. In radiation therapy, ionized radiation with treatment by photons (X-rays) along with other way of cancer manage-

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

ment like chemotherapy, surgery etc. is carried out. Medical imaging and fusion plays a significant role in aiding medical physicist or radiation oncologist for accurate diagnosis and treatment management of the cancer patients. Radiation therapy involves localization of infected cells, planning of radiation doses and treatment analysis. It is very important that diagnosis of spread of tumor or infected cells should be very accurate for treatment using radiotherapy. The images of infected body organ are captured through plain radiographs (X-rays), Magnetic Resonance Imaging (MRI), computed tomography (CT), Positron Emission Tomography (PET), Single-Photon Emission Computed Tomography (SPECT), functional Magnetic Resonance Imaging (fMRI) etc. These images are very useful to delineate tumor or infected tissues. Medical image fusion works towards integration of the above modality images for exact localization and quantifying volume (spread) of infected tissues/cells and/or lymph nodes. This will facilitate radiation oncologist for better treatment planning such as maximizing radiation dose to infected cells/tumor and minimize the radiation dose to healthy surrounding tissues and organs at risk (OAR).

LITERATURE SURVEY

Multimodality medical image fusion is much developed and widely used process in clinical applications. The fusion process utilizes variety of image data sets and information representations. Many methods, techniques and algorithms have been proposed in literature for the process (James, & Dasarathy, 2014). Fusion methods are typically based on the process of combining complementary information between two or more image data sets. Medical image fusion process may be achieved in spatial and frequency domain. It is also categorized into pixel level fusion, feature level fusion and decision level fusion. Pixel level fusion is spatial domain fusion which is much simpler and is computationally efficient; but it fails to provide enough contrast (Yang, Huang & Gao, 2014). The other popular spatial domain techniques are intensity-hue-saturation (IHS), principal component analysis (PCA), and the Brovey transform. These techniques are most commonly used in fusion of remote sensing images. These methods provide better spatial representation but carries less spectral representation (Pradhan, King, Younan, & Holcomb, 2006). Feature level fusion is frequency domain technique which utilizes extracting significant features and merging them with certain fusion criteria in frequency domain. Decision level fusion uses multi-sensor data for combining relevant information (Shivappa, Rao, & Trivedi, 2008).

Medical Image fusion is also experimented using artificial neural network (Li, Kwok, & Wang, 2002), fuzzy logic (Singh, Raj, Kaur, & Meitzler, 2004), neuro-fuzzy approach (Das, & Kundu, 2013). Researcher also combined CT-PET images (Deantonio et al., 2008; Grosu, Weber, & Franz, 2005) and MRI-PET images (Nagamachi et al., 2013) which helped physicians in nuclear medicine to get better activity visualization of infected volume. Now a day, it is used very effectively in diagnosis and treatment of cancer patients.

Recently many researchers have developed a large number of fusion techniques based on multi-resolution approach. The multi-resolution analysis is popularly used as an aid to extract features from an image; thus it proved to be a powerful tool for analyzing medical images. Multi-resolution analysis involves pyramid transform and wavelet transform. The methods of pyramid transform are Laplacian pyramid (Burt, & Adelson, 1983), contrast pyramid (Toet, van Ruyven & Valeton, 1989), gradient pyramid (Burt, & Kolczynski, 1993), and the morphological pyramid (Matsopoulos, & Marshall, 1995). The other multiresolution approach is the wavelet based methods in which multi-resolution frequency sub-band coefficients are derived using Discrete Wavelet Transform (DWT). DWT methods are much superior

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

over pyramid transforms due to its better spatio-spectral localization. These techniques provide better directionality, less visual artifacts and good signal-to-noise ratios (Mallat, 1989; Singh, & Khare, 2013).

There are many wavelet based methods found in literature developed for medical image fusion (Chipman, Orr, & Graham, 1995; Yang, Park, Huang, & Rao, 2010). The majority of fusion algorithms use maximum selection rule (Burt et al., 1993). The modulus maxima selection criteria applied on wavelet coefficients is given in Qu, Zhang, & Yan (2001). The wavelet coefficients are weighted to get fusion of PET and CT images presented in Cheng, He, & Lv (2008) whereas multiwavelet transform is used to get PET/CT fusion (Liu, Yang, & Sun, 2010). The multi-sensor medical image fusion of CT images with MRI images by wavelet transform is described in Li, Manjunath, & Mitra (1995). The decomposition of anatomical and functional images using Integer Wavelet Transform and combined them using Neuro-Fuzzy approach is presented by Kavitha et al. (2010).

In this paper, frequency domain multimodality medical image fusion techniques are discussed. The different sets of CT scan images and MRI images are used to merge relevant detailed medical information in a frequency domain. Different fusion rules are used for combining both set of images. The Discrete Fourier Transform (DFT), Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), Dual Tree Complex Wavelet Transform (DT-CWT), Daubechies Complex Wavelet Transform (DCWT) and M-Band Wavelet Transform (M-Band WT) are used to represent both set of CT and MRI images into frequency domain and fusion is achieved by applying appropriate fusion rules in frequency domain. The resultant or fused image is reconstructed by applying inverse transform on the planes of modified/selected coefficients.

MULTIMODALITY MEDICAL IMAGE FUSION SCHEME

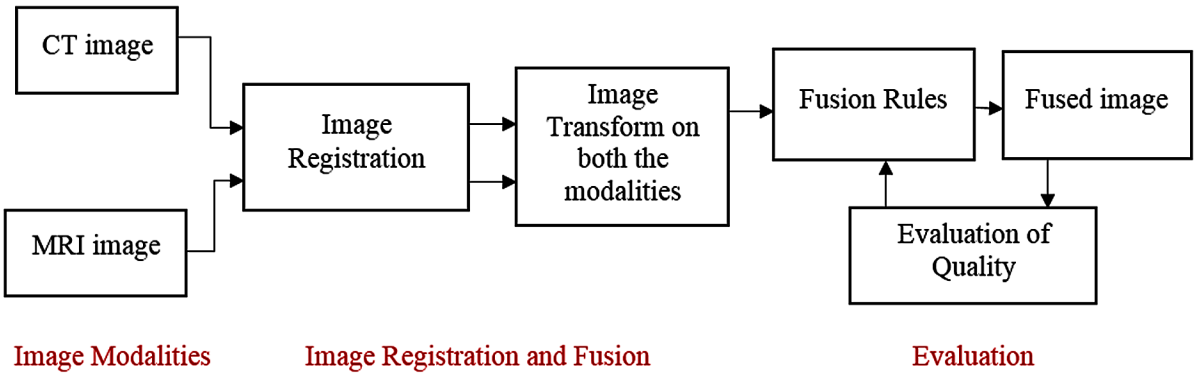
Multimodality Medical image fusion involves basic three steps namely image modalities, image registration and fusion, and evaluation as shown in Figure 1 (Chavan, & Talbar, 2014). Various modality images are applied as source images. These source images should be registered or aligned by suitable registration algorithm. The technique presented in this paper uses feature based fusion in which the features of source images are extracted using appropriate transform. These features are then merged by applying fusion rules on frequency domain coefficients. The evaluation step is adaptive and use feedback for modification of fusion rules. It also ensures better efficiency of the algorithm. The fused image must be a better quality image for visual interpretation and can be used for further processing if required. It should be free from noise and visual artifacts. It is also expected that the process should only combine complementary information from both the source images and it should not add any false/irrelevant information.

Image Modalities

Medical physicists, radiation oncologists, physicians in radio diagnostic and nuclear medicine are the important health professionals in cancer treatment. They use various imaging modalities like X-rays, ultrasound images (USG), computed tomography (CT), functional MRI (fMRI), magnetic resonance imaging (MRI), single-photon emission computed tomography (SPECT) and positron emission tomography (PET) separately or together for diagnosis and treatment planning. It is highly recommended to examine images from various modalities for accurate diagnosis as well as appropriate treatment plan (Kessler, & Li, 2001). Physicians chooses most appropriate image modalities for clinical examination.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 1. Steps in medical image fusion process



However, single medial image modality may not be sufficient to provide all necessary information. Therefore, physicians may use various modalities to arrive at final diagnosis. Thus, medical image fusion will definitely useful to these medical practitioners in this process of diagnosis and treatment.

In image acquisition process, the scanner systems play a key role to capture necessary information from human body. The captured images carry relevant information depending upon the modality. Some manufacturers developed acquisition devices which capture images with the help of two modalities. But such machines are extremely expensive. Therefore it is always preferred to have single modality imaging devices and with the help of fusion algorithms, one can combine important and most relevant image details in single image.

Medical Image Registration and Fusion

The registration and fusion helps to combine relevant and complementary information within different modalities which will be advantageous to physicians in cancer patient management. Medical images used in the process of registration and fusion are taken from different modalities. Therefore these images differ in position i.e. localization, orientation, and resolution (size and pixel representations); but these images are rich enough to provide relevant and complimentary information. These images should be properly aligned i.e. registered with each other so that medical image fusion can be effectively achieved to get better visual perception. The effectiveness and efficiency of the fusion algorithms are also depend upon the quality of registration process.

Image registration is the process to bring second image i.e. target image in alignment with first image i.e. source image. The process involves translation by ‘a’ unit in x-direction and ‘b’ unit in y-direction provided a, b are integers, rotation in clockwise or anti-clockwise direction by ‘θ’ degree and scaling of image (zooming/shrinking) to bring it equal to size of source image without losing important details. It is also called geometric transformation. It helps in bringing each and every pixel of both images properly aligned. Thus, the region of interest (ROI) i.e. tumor or region of infected tissues will be aligned in all the images after proper registration. Image Registration is classified into intra-subject registration where different views are used as a cue, inter-subject registration in which same organ of different persons is used and serial registration which helps to assess the changes in ROI over a stipulated period (Kessler, & Li, 2001).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Transforms and Fusion Rules

After registration, both the source images i.e. CT and MRI images are converted into frequency domain by applying one of the transforms. The transform coefficients of the source images are fused using appropriate set of fusion rules in frequency domain. The inverse transform is applied to get resultant image which is called fused image. The process is multimodality medical image fusion. The transforms used in this paper are Discrete Fourier Transform (DFT), Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), Dual Tree Complex Wavelet Transform (DT-CWT), Daubechies Complex Wavelet Transform (DCWT) and M-Band Wavelet Transform (M-Band WT).

Discrete Fourier Transform (DFT)

The Discrete Fourier Transform (DFT) is orthogonal and separable transform (Gonzalez, & Woods, 2003). The image is transformed into Fourier domain using Equation (1), while inverse Fourier transform of an image is calculated using Equation (2):

$$F_{DFT}\left(u,v\right)=\frac{1}{M}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f_s\left(x,y\right)e^{\frac{-j2\pi u}{M}x+\frac{-j2\pi v}{N}y} \tag{1}$$

$$f_s\left(x,y\right)=\frac{1}{N}\sum_{u=0}^{M-1}\sum_{v=0}^{N-1}F_{DFT}\left(u,v\right)e^{\frac{j2\pi u}{M}x+\frac{j2\pi v}{N}y} \tag{2}$$

where, $f_s\left(x,y\right)$ is original image of size MxN and $F_{DFT}\left(u,v\right)$ is Fourier transform of $f_s\left(x,y\right)$. The Fourier coefficients of source images are complex representations.

The first coefficient i.e. top left most coefficient is called DC coefficient. These DC coefficients of both the images are averaged in both real and imaginary plane and maximum valued remaining coefficients i.e. AC coefficients from one of the source images is selected to form transformation planes. The inverse discrete Fourier transform is applied on transformation planes to achieve fused image.

Discrete Cosine Transform (DCT)

The discrete cosine transform (DCT) has very high energy compaction ability and it is also computationally efficient. The block based DCT provides very good energy compaction preserving local information also. The source images are first divided into 8x8 non-overlapping blocks and two dimensional discrete cosine transform (2D-DCT) is applied on every block. Equation (3) and Equation (4) are used to compute 2D-DCT and two dimensional inverse discrete cosine transform (2D-IDCT) of the images. The energy of image gets concentrated to top left corner of every block after 2D-DCT. These few coefficients in left top corner of every block are sufficient to represent entire image information (Bhattacharya, Dey, & Chaudhuri, 2012; Gonzalez, & Woods, 2003; Soman, Ghorpade, Sonone, & Chavan, 2012). The first coefficient is called DC coefficient while all other coefficients are called AC coefficients:

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

$$F_{DCT}\left(u,v\right)=a_u a_v \sum_{i=0}^{M-1}\sum_{j=0}^{N-1} f_s\left(i,j\right)\cos\frac{\left(2i+1\right)\pi u}{2M}\cos\frac{\left(2j+1\right)\pi v}{2N} \tag{3}$$

$$f_s\left(i,j\right)=\sqrt{\frac{2}{M}}\sqrt{\frac{2}{N}}\sum_{u=0}^{M-1}\sum_{v=0}^{N-1} a_u a_v F_{DCT}\left(u,v\right)\cos\frac{\left(2i+1\right)\pi u}{2M}\cos\frac{\left(2j+1\right)\pi v}{2N} \tag{4}$$

where, $F_{DCT}\left(u,v\right)$ is 2D-DCT of source image $f_s\left(i,j\right)$ of size MxN and:

$$a_u = a_v = \begin{cases} \sqrt{\frac{1}{M}} & ;u=v=0 \\ \sqrt{\frac{1}{N}} & ;u\neq v\neq 0 \end{cases}$$

The DC coefficients of each block of both the images are averaged and maximum valued AC coefficient from one of the source images is selected to form transformation plane. The 2D-IDCT is applied on transformation plane to get fused image.

Wavelet Transform (DWT)

Discrete Wavelet Transform (DWT)

The multi-resolution analysis (MRA) property of Discrete Wavelet Transform (DWT) helps in decomposition of medical images into different frequency bands (Dey, N. et al. 2014; Mallat, 1989; Vetterli, & Herley, 1992). The wavelet transform is given by Equation (5):

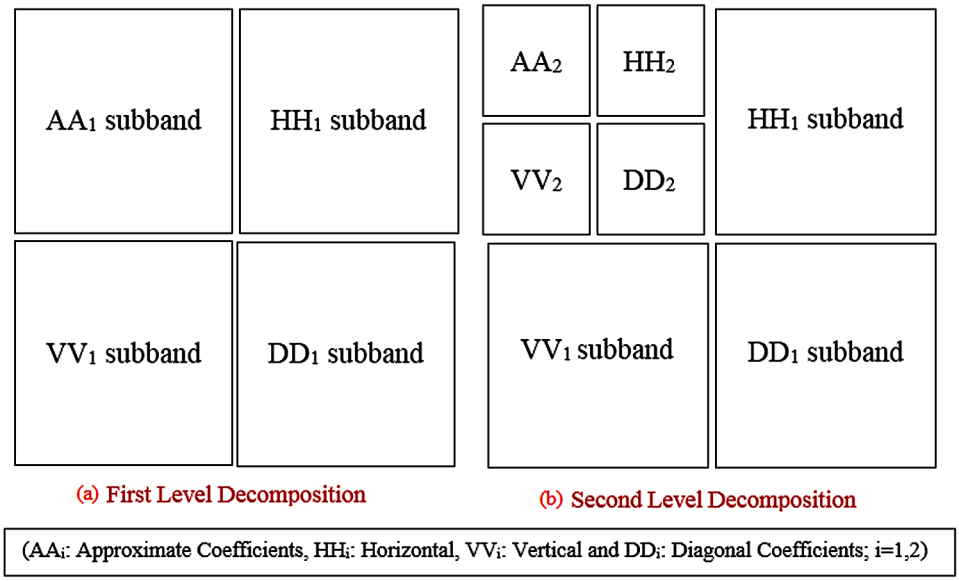
$$\psi_{\rho,x0}\left(x\right)=\frac{1}{\sqrt{\left|\rho\right|}}\psi\left(\frac{x-x0}{\rho}\right),\rho,x0,\in R,\rho\neq 0 \tag{5}$$

where, ρ stands for scaling which scales the information. The term $\psi\left(\frac{x-x0}{\rho}\right)$ is kernel of the wavelet as given in Equation (5). The parameter *x0* represents position of the kernel wavelet. Various values of scaling ρ results in different values of the wavelet function $\psi_{\rho,x0}\left(x\right)$ than that mother wavelet '$\psi\,(x)$'. Thus, the above wavelet functions have time widths adapted to their frequencies (Dey, N., 2012; Singh, & Khare, 2013).

The source medical images i.e. CT and MRI images are first decomposed using DWT. The first level decomposition sub-bands and second level decomposition sub-bands are as shown in Figure 2. Thus various levels of decomposition can be achieved with repetitive application of two dimensional discrete wavelet transform (2D-DWT) on approximate sub-band (AA) in every decomposition level. DWT provides excellent spatio-frequency localization and it is helpful in energy concentration of image in approximate sub-band (AA). The detailed information like edges, sharpness etc. get concentrated in

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 2. Two-dimensional discrete wavelet transform decomposition (a) First level decomposition (b) Second level decomposition



remaining three sub-bands viz. horizontal (HH), vertical (VV) and diagonal (DD) sub-bands. The edge information provided by DWT is in three orientations i.e. 0^0, 90^0, and $\pm 45^0$. This energy re-orientation in wavelet sub-bands provides excellent feasibility to combine images effectively.

When images are merged in wavelet space, different frequency sub-bands are fused differently using appropriate fusion rules. The DWT coefficients in approximate (AA) plane are averaged; while maximum valued coefficients in high frequency sub-bands i.e. HH, VV, DD planes are selected from both the modalities for combining them and composite plane is created. The inverse discrete wavelet transform (IDWT) is applied on composite plane to generate fused image.

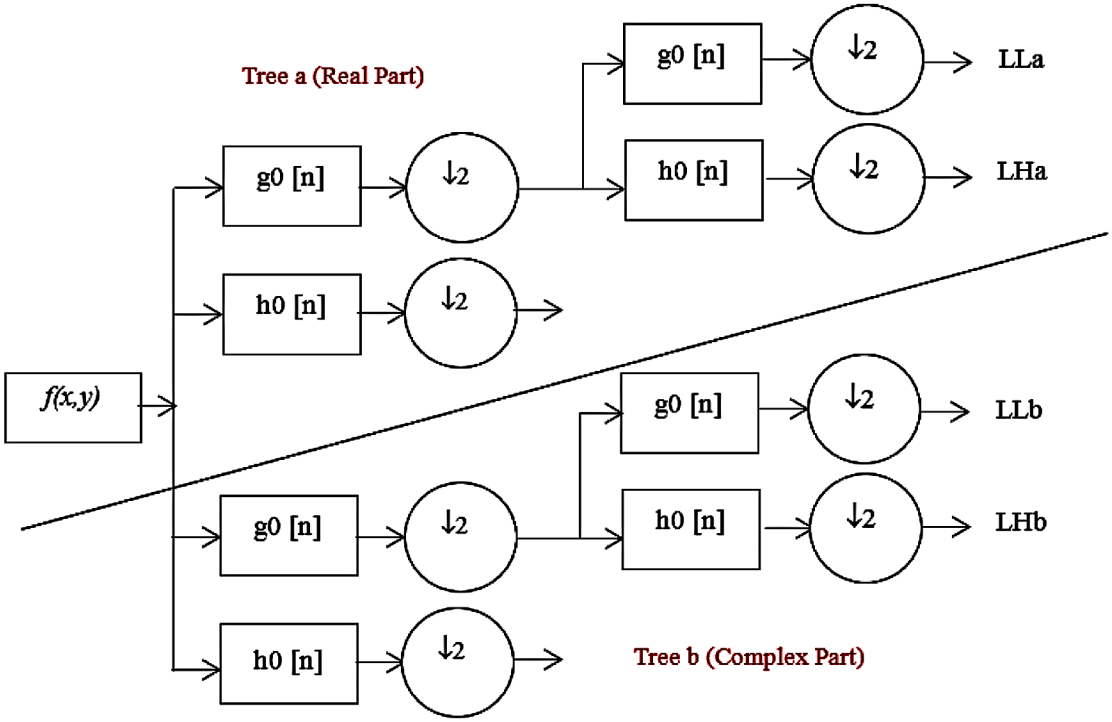
Dual Tree Complex Wavelet Transform (DT-CWT)

The DT-CWT is extension to the standard discrete wavelet transform (DWT) which is represented as complex valued wavelet transform. It is a two-dimensional two-tree wavelet transform which provides multi-resolution, sparse representation and useful characterization of the structure of an image. The DT-CWT uses two parallel wavelet filter banks in two tree structures (Selesnick, Baraniuk, & Kingsbury, 2005). One set of filter coefficients of first tree (Tree a) are treated as the real part (Xi); while another set of coefficients in second tree (tree b) are taken as complex part (Yi) as shown in Figure 3. The selection of filter coefficients are made in such way that aliasing will be avoided when down sampling is applied on coefficients after filtering. DT-CWT is not a true complex wavelet transform in true sense as all the coefficients are real; but it is used in complex representation ($Xi+jYi$) which provides phase information of decomposed images. The demerits like shift sensitivity and rotation variance of DWT can be ruled out by use of DT-CWT (Selesnick, Baraniuk, & Kingsbury, 2005; Singh, & Khare, 2014).

If the filters used in one tree are especially designed different from those in the other tree, it is possible for one DT-CWT to produce the real coefficients and the other the imaginary. The DT-CWT is

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 3. Dual tree complex wavelet transform (DT-CWT) decomposition of an image



f(x,y): **Original Image; g0 [n]: Low pass filter coefficients; h0[n]: High pass filter coefficients; LLa, LLb: Approximate Images of tree a and b; LHa, LHb: Horizontal details of image for tree a and b.**

implemented using two separate two-channel filter banks. The low-pass (scaling) filters and high-pass (wavelet) filters of one tree 'a',{g0, h0} must generate a scaling function and wavelet filters which are approximate Hilbert transforms of the scaling function and it is generated by low-pass and high-pass filter of other tree 'b',{g1, h1}. Therefore, the complex valued scaling function and wavelet function from the tree 'a' and tree 'b' are approximately analytic.

Two level decomposition using DT-CWT results into 14 sub bands; out of which 12 are high frequency bands and 2 are low frequency bands. DT-CWT provides edge information in 6 different orientations i.e. $\pm 15^0$, $\pm 45^0$, $\pm 75^0$ which is more than that of traditional DWT; but it is computationally expensive transform compared with DWT.

The DT-CWT approximate coefficients (AA) of magnitude plane are averaged to create magnitude plane of fused image and the maximum valued coefficients in the phase plane of detailed frequency sub-bands like HH, VV and, DD are selected to create phase plane. The inverse DT-CWT is applied on magnitude and phase plane to reconstruct fused image.

Daubechies Complex Wavelet Transform (DCWT)

The DCWT is actually true complex wavelet transform which provides better phase information and shift-invariance compared to dual tree complex wavelet transform (DT-CWT) (Clonda, Lina, & Goulard, 2004; Singh, & Khare, 2014). It is also computational efficient compared to DT-CWT. It is therefore most

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

appropriate frequency domain representation for multimodality medical image fusion. The Daubechies wavelet function is given by Equation (6) and its scaling function is given by Equation (7):

$$\zeta(x) = 2\sum_{i}(-1)^{i}\overline{\alpha_{1-i}}s(2x-i) \tag{6}$$

$$s(m) = 2\sum_{i}\alpha_{i}s(2m-i) \tag{7}$$

where, $\zeta(x)$ is mother wavelet and s (m) is the scaling function of DCWT. α_i are considered the real coefficients and sum of all coefficients (α_i) is equal to one (Clonda, Lina, & Goulard, 2004; Singh, & Khare, 2014).

M-Band Wavelet Transform (M-Band WT)

Two-band orthogonal wavelet has many drawbacks such as it has non-symmetric orthogonal coefficients. There is lack of freedom to select orientations i.e. directionality etc. Therefore, other alternative wavelet transforms like M-band wavelet transforms are devised. The high spectral information is extracted using M-band wavelets which results into better spatial as well as spectral representation. The source images are decomposed in multi-scale and multi-directional representations using M-Band wavelet transform. It will help to preserve detailed information from source images and thus provides a very good quality of multimodality medical image fusion.

The M-band discrete orthogonal wavelet is also constructed in same manner as 2-band. It consists of scaling function i.e. father wavelet ϕ(x) and wavelet function i.e. mother wavelet ψ(x) which are given in Equations (8) and (9) respectively (Zou & Tewfik, 1992). The single level decomposition of M-band wavelet (M=3, L=1) is represented in Figure 4:

$$\phi(x) = \sum\alpha_{i}\phi(Mx-i) \tag{8}$$

where, M ≥ 2 indicating scale and α_i are scaling coefficients:

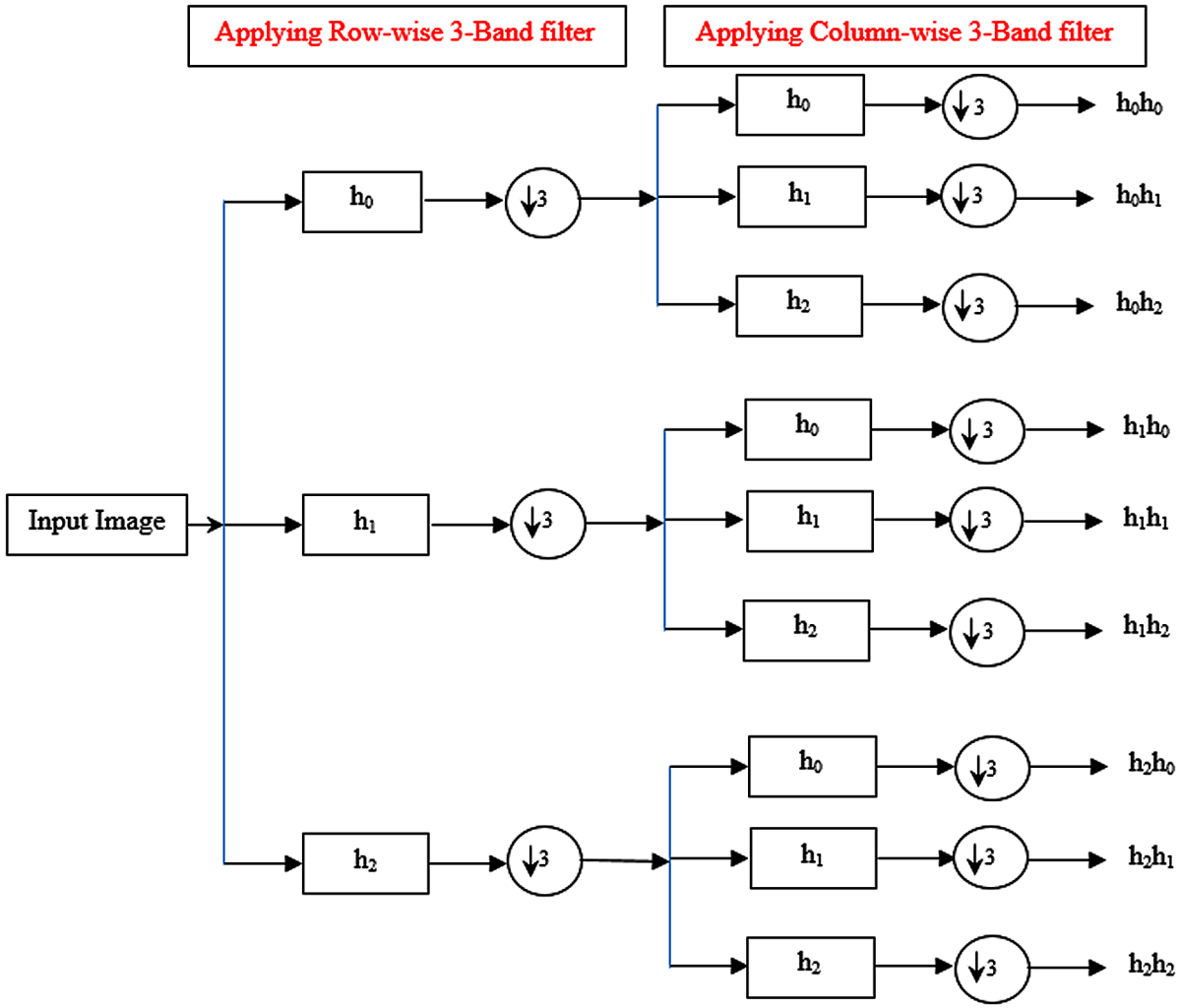
$$\psi_{k}(x) = \sum\beta_{i}^{k}\phi(Mx-i) \tag{9}$$

where, β_i^k are (M-1) wavelet coefficients.

The M-Band wavelet transform is separable and orthogonal and has a perfect reconstruction quadrature mirror filter (QMF) representation. The M-Band one dimensional wavelet filters coefficients are used to generate 2D separable M^2 band transform which are represented as $\psi_{i,j}$ for i,j=1,2,3 with M=3 and ijth resolution can be obtained using filter $H_{i,j} = \psi_{i,j}\psi_{i,j}^{*}$ for i,j=1,2,3 with M=3. The source images are first decomposed into M^2 sub-bands by applying orthogonal and linear phase M-band Wavelet transform. It results into sub-bands which carries information at different resolution and directions (Aydin, Yemez, Anarim, & Sankur, 1996). When the M-band wavelet filters coefficients are applied on a source image, the image gets decomposed into one approximation sub-band and (M^2-1) detailed sub-bands. These (M^2-1) detailed sub-bands of decomposed image provide better edge information at different

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 4. M-band wavelet decomposition for M=3 and L=1



h_i, i=0,1,2: 3-Band Filter Coefficients; hihj, i,j=0,1,2: 3-band filtered planes; h_0h_0: Approximate Plane and remaining (M^2-1): Detailed planes

resolution and different orientation. M-band decomposition provides frequency representation in both logarithmic and linear scale. These merits of M-Band wavelet transform make it a better algorithm or approach for feature extraction in the multimodality image fusion. The demerit with M-band wavelet is its computational expensiveness.

The fusion rule applied in approximate plane is averaging where average of the wavelet coefficients from both the source images is taken in resultant plane; while maximum selection rule is applied in remaining (M^2-1) sub-bands.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

PERFORMANCE METRICS FOR COMPARISON

The effectiveness and efficiency of the algorithm can be justified after comparison of the results. The performance evaluation is subjective as well as objective analysis. The subjective evaluation is based on visual interpretation of fused images by radiologist. The objective evaluation is parameter based evaluation in which performance parameters are computed. The various performance parameters used in this paper are Mean, Standard Deviation, Entropy, Fusion Factor, Correlation Coefficients and Cross Entropy. The source CT scan and MRI images and their fused medical image are used to estimate values of the performance parameters. Higher or lower value of particular metrics proves the effectiveness of that algorithm.

Mean (μ_{fus})

Average pixel value of the image is called its mean. It is computed using Equation (10):

$$\mu_{fus} = \frac{1}{MxN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} f_{fus}(i,j) \tag{10}$$

Higher value of 'mean' $\left(\mu_{fus}\right)$ is nothing but better quality of the fused image.

Standard Deviation (σ_{fus})

Standard deviation provides the distribution of information content from its mean value and can be effective to decide the quality of fused image. It is calculted by using Equation (11):

$$\sigma_{fus} = \left(\frac{1}{MxN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \left(f_{fus}(i,j) - \mu_{fus} \right)^2 \right)^{1/2} \tag{11}$$

where μ_{fus} is mean and σ_{fus} is standard deviation of fused image $f_{fus}(i,j)$. Higher value of standard deviation means the quality of fused image is better compared to original images.

Entropy (En)

Image information content can be better represented by its entropy. Higher entropy value of images provides more information than lower entropy value of images. Entropy is calculated using Equation (12):

$$En = -\sum_{i=0}^{M-1} P_i \log_2 P_i \tag{12}$$

where, P_i is probability distribution of gray values in the fused image.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Fusion Factor (FusFac)

Fusion factor (*FusFac*) is based on the mutual information between original images (CT and MRI) and resultant fused image. Mutual information is one of the measures which indicates how much information of one image is present in another image. Fusion factor which is given by Equation (13) is sum of mutual information of original images and fused images. Higher value of fusion factor *(FusFac)* indicates better quality of resultant or fused image:

$$FusFac = MI_{1,fus} + MI_{2,fus} \tag{13}$$

where $MI_{1,fus}$ provides mutual information index for CT image and fused image, $MI_{2,fus}$ provides mutual information index for MRI image and fused image.

The Mutual information $MI_{A,B}$ between A and B images is given by Equation (14):

$$MI_{A,B} = \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} P_{A,B}(x,y) \log \frac{P_{A,B}(x,y)}{P_A(x) P_B(y)} \tag{14}$$

where, $P_{A,B}$ is joint probability density function of two images A and B whereas P_A and P_B are probability density function of images A and B respectively.

Correlation Coefficients ($\mathcal{R}corr$)

The correlation coefficient $(\mathcal{R}corr)$ is a measure of closeness between source images and fused image. Correlation coefficient $(\mathcal{R}corr)$ is calculated using Equation (15). The fused image is of better quality if the correlation coefficient is close to value 1:

$$\mathcal{R}_{corr}(x,y) = \frac{2\sum_{i=0}^{M-1}\sum_{j=0}^{N-1} f_o(i,j).f_{fus}(i,j)}{\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}\left|f_o(i,j)\right|^2 + \sum_{i=0}^{M-1}\sum_{j=0}^{N-1}\left|f_{fus}(i,j)\right|^2} \tag{15}$$

where, $f_o(i,j)$ is a source image of size MxN and $f_{fus}(i,j)$ is a fused image.

Cross Entropy (CEn)

The cross entropy is a measure of dis-similarity between the source images and the fused image. It is calculated using Equation (16):

$$CEn = \sum_{i=0}^{N-1} f_{pi} \log_2 \frac{f_{pi}}{g_{pi}} \tag{16}$$

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

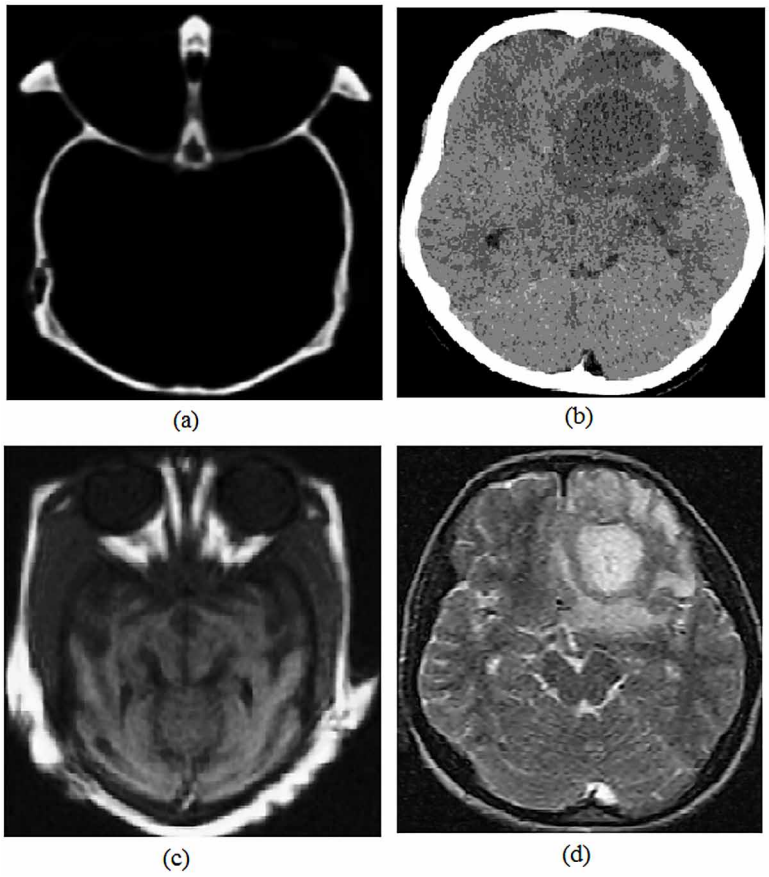
where f_{pi} and g_{pi} are distribution of the gray values in source images and fused image respectively. The quality of fused image is better if cross entropy is smaller in value.

RESULTS AND DISCUSSIONS

The multimodality medical image fusion is experimented with four different sample sets of medical images. The single set includes CT scan image slice and MRI image slice of a brain of the same patient. The resolution of CT scan images is 512 x 512, 16 bits per pixel and MRI images is 512 x 320, 16 bits per pixel. Source CT scan image and MRI image for the sample image slices from medical image data set 1 and set 2 are as shown in Figure 5.

First, both CT and MRI images from one set are registered using appropriate geometric transformation. The registration process involves translation or shifting, scaling and rotation transformation to bring both the images in pixel alignment. Then different transforms like discrete Fourier transform (DFT), discrete cosine transform (DCT), discrete wavelet transform (DWT), dual tree complex wavelet transform (DT-CWT), Daubechies complex wavelet transform (DCWT), and M-Band wavelet transform

Figure 5. Two original sample set of source images: (a) CT scan image of set 1; (b) CT scan image of set 2; (c) MRI image of set 1; and (d) MRI image of set 2



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

(M-band WT) are applied on both modality images separately. The DFT and all wavelet transforms are applied on entire images while the images are first subdivided into 8x8 non-overlapping blocks before applying DCT. The images are decomposed in various levels of decomposition in wavelet transform. The frequency domain coefficients are then fused using various fusion rules depending on transform used. The selection of the fusion rules depends on the process of extracting information from both the images.

Objective Evaluation

The estimation of visual quality of the resultant image after fusion algorithms using various performance metrics like mean (μ_{fus}), standard deviation (σ_{fus}), entropy (En), fusion factor (FusFac), correlation coefficient $\left(\mathcal{R}corr\right)$, and cross entropy (CEn) is presented in this section as an objective analysis. The m-band wavelet fusion and Daubechies complex wavelet transform fusion are compared with fusion techniques using dual tree complex wavelet transform, discrete wavelet transform, discrete Fourier transform, and discrete cosine transform. The performance metrics are such that higher values of mean, standard deviation, entropy, fusion factor reveals better quality of fused image; whereas correlation coefficient must be close to one and cross entropy should be comparatively low to declare that the fused image is of good quality. It is objective evaluation process. The performance evaluation with fusion metrics are tabulated in Table 1 for four different sets of medical images. The values of mean, standard deviation, entropy, fusion factor are increasing for fused images using DFT, DCT, DWT, DT-CWT, DCWT and M-Band WT respectively. These values are higher for M-band wavelet transform and DCWT. The parameter correlation coefficient is approaching close to 1 for M-band WT and DCWT which is sign of better quality of fused images. The parameter cross entropy is decreasing from fused image using DFT, DCT, DWT, DT-CWT, DCWT and M-band WT respectively. Thus it is observed that the M-band wavelet transform and Daubechies complex wavelet transform gives higher values of performance parameters for mean, standard deviation, entropy, fusion factor. The value of correlation coefficient is very close to 1 and the value of cross entropy is very small for fused images using M-band wavelet transform and Daubechies complex wavelet transform. Therefore, M-band wavelet transform and Daubechies complex wavelet transform are the best transforms for multimodality medical image fusion.

The original source images i.e. CT and MRI are also compared with fused images for parameters like mean, standard deviation and entropy. The comparison results between original images and fused images for two sets of medical images (sample set 1 and set 2) are tabulated in Table 2. It is observed that the value of parameter 'mean' for fused images using DFT, DCT, DWT, DT-CWT, DCWT and M-band WT is higher than that of CT and MRI for both set of images. The values of parameters 'standard deviation' and 'entropy' for fused images with DFT and DCT are higher than that of CT and less than that of MRI; whereas these values are higher for fused images using DWT, DT-CWT, DCWT and M-band WT than that of both CT and MRI images. Thus it proves that fused images are of better quality than that of original source images.

The comparative fusion results of wavelet domain fusion with various levels of decomposition applied on sample set 3 are presented in Table 3. The source images are decomposed for four levels of decompositions. The fused images is generated after every level of decomposition. The results show that as decomposition levels are increased, the better quality of fused image is achieved in almost all wavelet transforms based on performance parameter values.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 1. Comparison of fusion algorithms for four sets of medical image sample data based on performance parameters

Sample Images	Method	Mean (μ_{fus})	Standard Deviation (σ_{fus})	Entropy (*En*)	Fusion Factor (*FusFac*)	Cross Correlation (Rcorr)	Cross Entropy (*CEn*)
Set 1	DFT	55.9310	7.0402	5.0195	1.1319	0.62793	3.5519
	DCT	57.6925	7.8578	6.0122	3.1307	0.65815	3.4664
	DWT	60.1721	8.6738	6.8446	4.5330	0.64642	3.0793
	DT-CWT	60.1527	8.6931	6.8934	4.8981	0.64381	2.6521
	DCWT	61.8217	8.6389	6.8754	4.8924	0.64107	2.3268
	M-Band WT	61.8273	8.6427	6.8781	4.8968	0.64110	2.1195
Set 2	DFT	67.6971	11.7915	6.3078	2.4571	0.90340	2.5124
	DCT	67.7336	11.6712	6.4868	4.9438	0.97711	1.4590
	DWT	79.6085	13.6862	6.6100	5.0940	0.98660	1.3486
	DT-CWT	79.5569	13.6024	6.6947	5.3004	0.98707	1.1230
	DCWT	79.8142	13.6118	6.7214	5.9560	0.99113	1.0129
	M-Band WT	79.8216	13.6154	6.7301	5.9623	0.99224	1.0017
Set 3	DFT	51.0891	10.7808	6.3300	2.2117	0.98643	1.9773
	DCT	54.9052	10.6927	6.5539	3.7494	0.99647	1.3331
	DWT	72.4438	11.0168	6.7099	4.1663	0.99883	1.2396
	DT-CWT	72.3521	12.9785	6.7543	5.2975	0.99934	1.0583
	DCWT	72.4513	12.9876	6.8428	5.3014	0.99941	1.0019
	M-Band WT	72.4589	12.9721	6.8526	5.3125	0.99982	0.9951
Set 4	DFT	54.2142	11.3005	5.1375	2.5544	0.9701	2.2824
	DCT	55.2314	11.4508	6.0390	3.0276	0.9804	1.4996
	DWT	70.6972	13.941	6.4344	4.7435	0.99839	1.3175
	DT-CWT	70.6335	13.9022	6.4583	5.7278	0.99952	1.2079
	DCWT	70.6917	13.9901	6.5917	5.8160	0.99987	1.1103
	M-Band WT	70.7108	13.9864	6.7240	5.8823	0.99993	1.0432

Subjective Evaluation

The resultant fused images from various transform domain fusion algorithms are also evaluated based on subjective analysis. The original CT and MRI images along with fused images using DFT, DCT, DWT, DT-CWT, DCWT and M-band WT for sample set 1 and sample set 2 are given in Figure 6 and Figure 7 respectively. Fused images of sample set 3 are presented in first row of Figure 8 and fused images of sample set 4 are presented in second row of Figure 8. The subjective analysis of the sample resultant images shows that M-Band WT, DCWT, and DT-CWT techniques provide better subjective quality images for visual perception compared to fused images using DWT, DCT, and DFT in almost all sets of medical images. They almost represent most of the anatomical information from both the modalities. The fusion using DFT in all algorithms provides inferior quality of image and fused image using DWT in sample set 2 and 3 are also visually inferior.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 2. Comparison of fused image with source CT and MRI images based on performance parameters like mean, standard deviation, and entropy

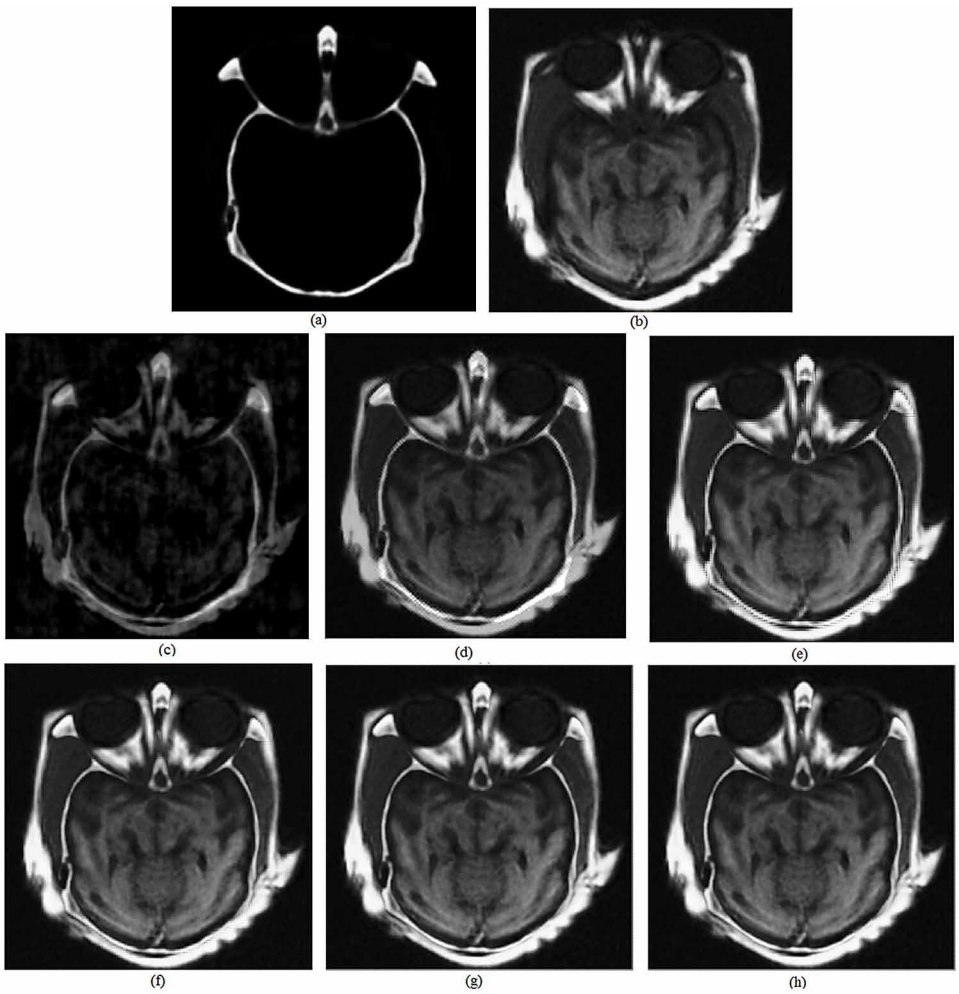
		Sample Set 1			Sample Set 2		
Method	**Image**	**Mean (μ_{fus})**	**Standard Deviation (σ_{fus})**	**Entropy *(En)***	**Mean (μ_{fus})**	**Standard Deviation (σ_{fus})**	**Entropy *(En)***
Source	CT	11.2461	4.7330	1.8352	66.8186	10.7907	4.6158
Source	MRI	54.5858	8.4756	6.7154	61.1148	12.7765	6.0218
DFT	Fused	55.9310	7.0402	5.0195	67.6971	11.7915	6.3078
DCT	Fused	57.6925	7.8578	6.0122	67.7336	11.6712	6.4868
DWT	Fused	60.1721	8.6738	6.8446	79.6085	13.6862	6.6100
DT-CWT	Fused	60.1527	8.6931	6.8934	79.5569	13.6024	6.3947
DCWT	Fused	61.8217	8.6389	6.8754	79.8142	13.6118	6.7214
M-Band WT	Fused	61.8273	8.6427	6.8781	79.8216	13.6154	6.7301

Table 3. Comparison of wavelet transform applied on sample set 3 for different levels of decompositions

Transform	**Level**	**Mean μ_{fus}**	**Standard Deviation σ_{fus}**	**Entropy En**	**Fusion Factor FusFac**	**Correlation Coefficient $\mathcal{R}corr$**	**Cross Entropy CEn**
DWT	I	72.4102	11.0015	6.6914	4.1005	0.99524	1.2405
	II	72.4438	11.0168	6.7099	4.1663	0.99883	1.2396
	III	72.4681	11.1200	6.7185	4.1980	0.99889	1.2188
	IV	72.4982	11.1358	6.7267	4.1996	0.99901	1.2161
DT-CWT	I	72.1598	12.8940	6.7102	5.2891	0.99821	1.4630
	II	72.3521	12.9785	6.7543	5.2975	0.99934	1.0583
	III	72.3894	12.9856	6.7598	5.3084	0.99986	1.0349
	IV	72.3957	12.9893	6.7610	5.3357	0.99991	1.0154
DCWT	I	72.4489	12.9821	6.8400	5.2993	0.99933	1.0164
	II	72.4513	12.9876	6.8428	5.3014	0.99941	1.0019
	III	72.4580	12.9892	6.8434	5.3120	0.99956	1.0004
	IV	72.4599	12.9903	6.8491	5.3328	0.99972	0.9994
M-Band DWT	I	72.3467	12.9541	6.8401	5.3021	0.99945	0.9994
	II	72.4589	12.9721	6.8526	5.3125	0.99982	0.9951
	III	72.6540	12.9801	6.8539	5.3187	0.99989	0.9946
	IV	72.6985	12.9874	6.8584	5.3210	0.99992	0.9913

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 6. Fused images of sample set 1: (a) Original CT image; (b) Original MRI image; (c) Fusion result using DFT; (d) Fusion result using DCT; (e) Fusion result using DWT; (f) Fusion result using DT-CWT; (g) Fusion result using DCWT; (h) Fusion result using M-band wavelet

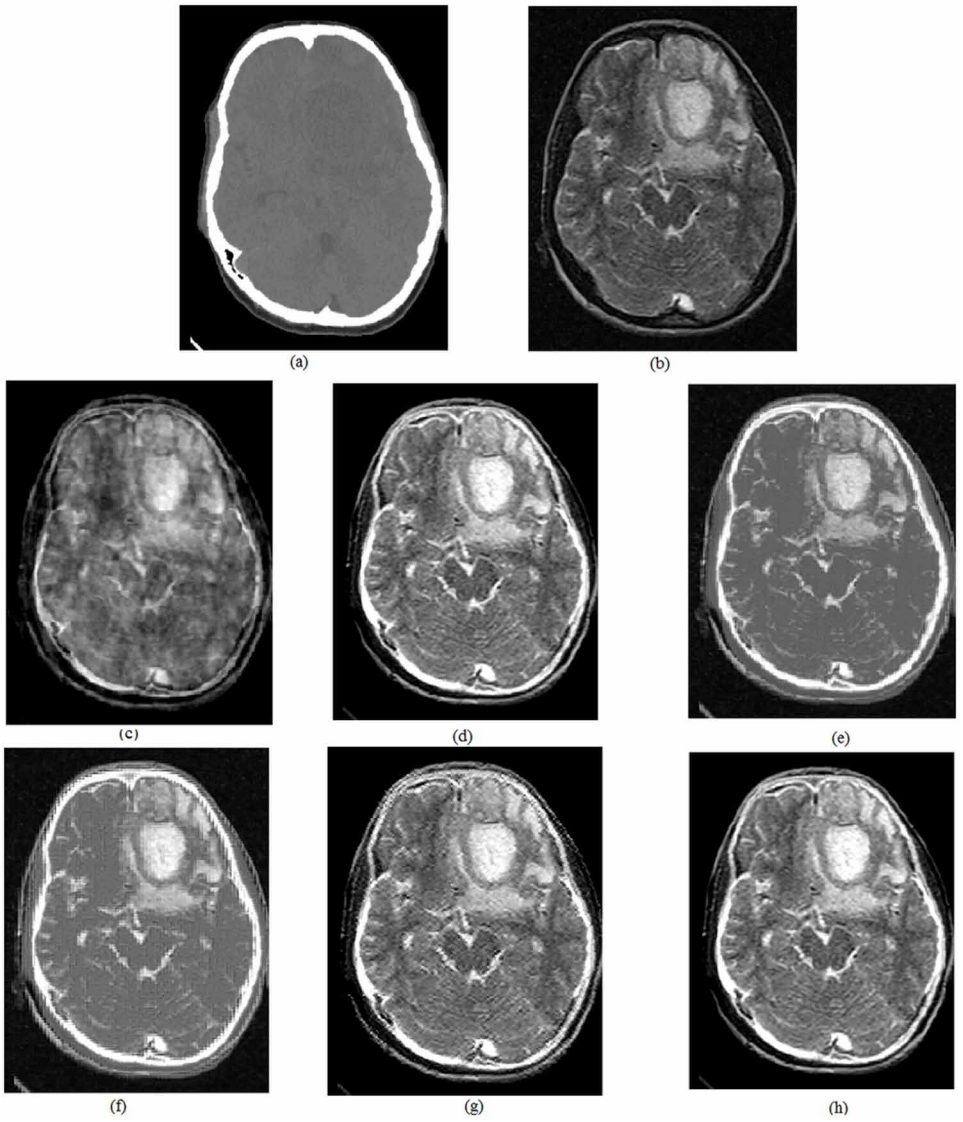


CONCLUSION

Modern cancer treatment uses multimodality medical image fusion effectively for diagnosis and treatment. The efficient fusion algorithms provide fused images from different modalities to diagnose and precisely delineate infected tissues with respect to tumor size, its spread, and its exact position. The fusion algorithms presented in this paper are feature based fusion algorithms. The features of source images are extracted using Discrete Fourier Transform (DFT), Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), Dual Tree Complex Wavelet Transform (DT-CWT), Daubecheis Complex Wavelet Transform (DCWT), and M-Band Wavelet Transform (M-Band WT). These features are fused using appropriate fusion rules in spectral domain. The M-Band Wavelet Transform and Daubechies Complex Wavelet Transform techniques are more effective based on objective evaluation process compared with other transforms. The subjective quality of the fused images using M-Band Wavelet, Daubechies Com-

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

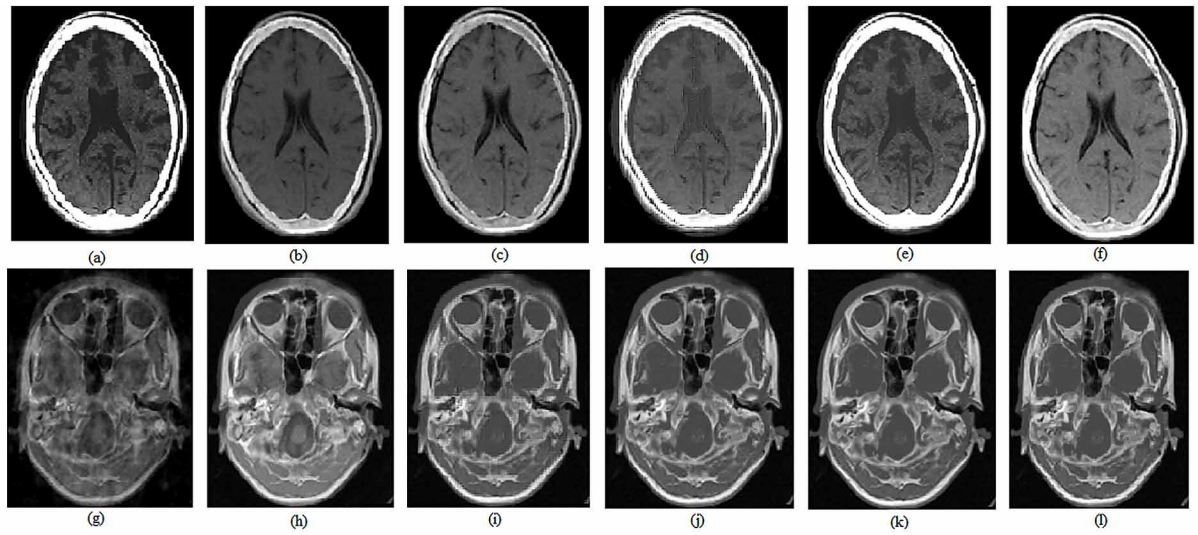
Figure 7. Fused images of sample set 2: (a) Original CT image; (b) Original MRI image; (c) Fusion result using DFT; (d) Fusion result using DCT; (e) Fusion result using DWT; (f) Fusion result using DT-CWT; (g) Fusion result using DCWT; (h) Fusion result using M-band wavelet



plex Wavelet and Dual Tree Complex Wavelet transforms are much better than that of technique using Discrete Fourier Transform, Discrete Cosine Transform and Discrete Wavelet Transform. The fused images using M-band WT, DCWT and DT-CWT provide more visual information without visual artifacts and these images are of great help to physicians for accurate diagnosis and better treatment planning in cancer patient management.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 8. Fused images using different transforms. First row: Fused images of sample set 3. Second row: Fused images of sample set 4 using (a), (g): DFT (b), (h): DCT (c), (i): DWT (d), (j): DT-CWT (e), (k): DCWT (f), (l): M-band wavelet.



ACKNOWLEDGMENT

We are very grateful to Ekdant Diagnostics Centre, Nanded, Maharashtra, India for providing us multimodality medical image data sets of same patients for our research work. We would also like to thank http://www.imagefusion.org from where we received some of the source images for testing of developed algorithms. We are also thankful to anonymous reviewers for their meticulous approach of review and constructive suggestions for betterment of this manuscript.

REFERENCES

Araki, T., Ikeda, N., Dey, N., Acharjee, S., Molinari, F., Saba, L., & Suri, J. S. et al. (2015). Shape-Based Approach for Coronary Calcium Lesion Volume Measurement on Intravascular Ultrasound Imaging and Its Association with Carotid Intima-Media Thickness. *Journal of Ultrasound in Medicine*, *34*(3), 469–482. doi:10.7863/ultra.34.3.469 PMID:25715368

Araki, T., Ikeda, N., Dey, N., Chakraborty, S., Saba, L., Kumar, D., & Suri, J. S. et al. (2015). A comparative approach of four different image registration techniques for quantitative assessment of coronary artery calcium lesions using intravascular ultrasound. *Computer Methods and Programs in Biomedicine*, *118*(2), 158–172. doi:10.1016/j.cmpb.2014.11.006 PMID:25523233

Araki, T., Ikeda, N., Molinari, F., Dey, N., Acharjee, S., Saba, L., & Suri, J. S. (2014). Link between automated coronary calcium volumes from intravascular ultrasound to automated carotid IMT from B-mode ultrasound in coronary artery disease population. *International Angiology*, *33*(4), 392–403. PMID:25056172

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Araki, T., Ikeda, N., Molinari, F., Dey, N., Acharjee, S., Saba, L., & Suri, J. S. et al. (2014). Effect of Geometric-Based Coronary Calcium Volume as a Feature Along with its Shape-Based Attributes for Cardiological Risk Prediction from Low Contrast Intravascular Ultrasound. *Journal of Medical Imaging and Health Informatics*, *4*(2), 255–261. doi:10.1166/jmihi.2014.1251

Aydin, T., Yemez, Y., Anarim, E., & Sankur, B. (1996). Multidirectional and multiscale Edge Detection via M-Band Wavelet Transform. *IEEE Transactions on Image Processing*, *5*(9), 1370–1377. doi:10.1109/83.535850 PMID:18285227

Bhattacharya, T., Dey, N., & Chaudhuri, B. S. R. (2012). A Session Based Multiple Image Hiding Technique using DWT&DCT. *International Journal of Computers and Applications*, *38*(5), 18–21. doi:10.5120/4604-6808

Burt, P., & Kolczynski, R. (1993). Enhanced image capture through fusion.*4th IEEE International Conference on Computer Vision*, 173–182.

Burt, P. J., & Adelson, E. H. (1983). The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, *31*(4), 532–540. doi:10.1109/TCOM.1983.1095851

Chavan, S. S., & Talbar, S. N. (2014). Multimodality Image Fusion in Frequency Domain for Radiation Therapy.*IEEE International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom)*, 174-178. doi:10.1109/MedCom.2014.7005998

Cheng, S. L., He, J. M., & Lv, Z. W. (2008). Medical image of PET/CT weighted fusion based on wavelet transform.*Second International Conference on Bioinformatics and Biomedical Engineering*, 2523–2525. doi:10.1109/ICBBE.2008.964

Chipman, L., Orr, T., & Graham, L.N. (1995). Wavelets and image fusion. *Wavelet Applications in Signal and Image Processing III*, SPIE-2569.

Clonda, D., Lina, J. M., & Goulard, B. (2004). Complex Daubechies wavelet: Properties and statistical image modelling. *Signal Processing*, *84*(1), 1–23. doi:10.1016/j.sigpro.2003.06.001

Das, S., & Kundu, M. K. (2013). A Neuro-Fuzzy Approach for Medical Image Fusion. *IEEE Transactions on Bio-Medical Engineering*, *60*(12), 3347–3353. doi:10.1109/TBME.2013.2282461 PMID:24058012

Deantonio, L., Beldi, D., Gambaro, G., Loi, G., Brambilla, M., Inglese, E., & Krengli, M. (2008). FDG-PET/CT imaging for staging and radiotherapy trearment planning of haed and neckcarcinoma. *Radiation Oncology, 3(29),* pages 06.

Depeursinge, A., Rodriguez, A. F., Ville, D. V. D., & Muller, H. (2014). Three-dimensional solid texture analysis in biomedical imaging: Review and opportunities. *Elsevier Medical Image Analysis*, *18*(1), 176–196. doi:10.1016/j.media.2013.10.005 PMID:24231667

Dey, N. (2012). DWT-DCT-SVD based intravascular ultrasound video watermarking.*World Congress on Information and Communication Technologies (WICT)*, 224 - 229. doi:10.1109/WICT.2012.6409079

Dey, N., Samanta, S., Chakraborty, S., Das, A., Chaudhuri, S. S., & Suri, J. S. (2014). Firefly Algorithm for Optimization of Scaling Factors during Embedding of Manifold Medical Information: An Application in Ophthalmology Imaging. *Journal of Medical Imaging and Health Informatics*, *4*(3), 384–394. doi:10.1166/jmihi.2014.1265

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Gonzalez, R. C., & Woods, R. E. (2003). Digital Image Processing. Pearson Education, Second Edition, Third Indian Reprint, India, pages-154 and 472.

Grosu, A. L., Weber, W. A., Franz, M., Stärk, S., Piert, M., Thamm, R., & Nieder, C. et al. (2005). Reirradiation of recurrent high-grade gliomas using amino acid PET (SPECT)/CT/MRI image fusion to determine gross tumor volume for stereotactic fractionated radiotherapy. *International Journal of Radiation Oncology, Biology, Physics*, *73*(2), 511–519. doi:10.1016/j.ijrobp.2005.01.056 PMID:16168843

Ikeda, N., Araki, T., Dey, N., Bose, S., Shafique, S., El-Baz, A., & Suri, J. S. et al. (2014). Automated and accurate carotid bulb detection, its verification and validation in low quality frozen frames and motion video. *International Angiology*, *33*(6), 573–589. PMID:24658129

Ikeda, N., Gupta, A., Dey, N., Bose, S., Shafique, S., Araki, T., & Suri, J. S. et al. (2015). Improved Correlation between Carotid and Coronary Atherosclerosis SYNTAX Score Using Automated Ultrasound Carotid Bulb Plaque IMT Measurement. *Ultrasound in Medicine & Biology*, *41*(5), 1247–1262. doi:10.1016/j.ultrasmedbio.2014.12.024 PMID:25638311

James, A. P., & Dasarathy, B. V. (2014). Medical image fusion: A survey of the state of the art. *Elsevier Information Fusion*, *19*, 4–19. doi:10.1016/j.inffus.2013.12.002

Kavitha, C. T., & Chellamuthu, C. (2010). Multimodal medical image fusion based on Integer Wavelet Transform and Neuro-Fuzzy.*IEEE International Conference on Signal and Image Processing (ICSIP)*, 296-300. doi:10.1109/ICSIP.2010.5697486

Kessler, M. L. (2006). Image registration and data fusion in radiation therapy. *The British Journal of Radiology*, *79*(1), S99–S108. doi:10.1259/bjr/70617164 PMID:16980689

Kessler, M. L., & Li, K. (2001). Image Fusion for Conformal Radiotherapy. The American Association of Physicists in Medicine, 1-12.

Li, H., Manjunath, B., & Mitra, S. (1995). Multisensor image fusion using the wavelet transform. *Graphical Models and Image Processing*, *57*(3), 235–245. doi:10.1006/gmip.1995.1022

Li, S., Kwok, J., & Wang, Y. (2002). Multifocus image fusion using artificial neural networks. *Pattern Recognition Letters*, *23*(8), 985–997. doi:10.1016/S0167-8655(02)00029-6

Liu, Y., Yang, J., & Sun, J. (2010). PET/CT medical image fusion algorithm based on multiwavelet transform. *IEEE 2nd International Conference on Advanced Computer Control (ICACC), 2,* 264-268.

Mallat, S. G. (1989). A Theory for Multiresolution Signal Decomposition: A Wavelet Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *11*(7), 674–693. doi:10.1109/34.192463

Matsopoulos, G. K., & Marshall, S. (1995). Application of morphological pyramids: Fusion of MR and CT phantoms. *Journal of Visual Communication and Image Representation*, *6*(2), 196–207. doi:10.1006/jvci.1995.1018

Nagamachi, S., Nishii, R., Wakamatsu, H., Mizutani, Y., Kiyohara, S., Fujita, S., & Kawai, K. et al. (2013). The usefulness of (18) F-FDG PET/MRI fusion image in diagnosing pancreatic tumor: Comparison with (18) F-FDG PET/CT. *Annals of Nuclear Medicine*, *27*(6), 554–563. doi:10.1007/s12149-013-0719-3 PMID:23580090

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Pradhan, P. S., King, R. L., Younan, N. H., & Holcomb, D. W. (2006). Estimation of the number of decomposition levels for a wavelet-based multiresolution multisensor image fusion. *IEEE Transactions on Geoscience and Remote Sensing*, *44*(12), 3674–3686. doi:10.1109/TGRS.2006.881758

Qu, G. H., Zhang, D. L., & Yan, P. F. (2001). Medical image fusion by wavelet transform modulus maxima. *Optics Express*, *9*(4), 184–190. doi:10.1364/OE.9.000184 PMID:19421288

Selesnick, I. W., Baraniuk, R. G., & Kingsbury, N. G. (2005). The dual-tree complex wavelet transform. *IEEE Signal Processing Magazine*, *22*(6), 123–151. doi:10.1109/MSP.2005.1550194

Shivappa, S. T., Rao, B. D., & Trivedi, M. M. (2008). An iterative decoding algorithm for fusion of multimodal information. *EURASIP Journal on Advances in Signal Processing*, *2008*(1), 478396. doi:10.1155/2008/478396

Singh, H., Raj, J., Kaur, G. & Meitzler T. (2004). Image Fusion using Fuzzy Logic and its applications, *IEEE Fuzzy Systems*, 337-340.

Singh, R., & Khare, A. (2013). Multiscale medical image fusion in wavelet domain. *TheScientificWorld-Journal*, ID521034. PMID:24453868

Singh, R., & Khare, A. (2014). Fusion of multimodal medical images using Daubechies complex wavelet transform – A multiresolution approach. *Elsevier Information Fusion*, *19*, 49–60. doi:10.1016/j.inffus.2012.09.005

Soman, S., Ghorpade, M., Sonone, V., & Chavan, S. (2012). Content Based Image Retrieval Using Advanced Color and Texture Features.[IJCA]. *International Journal of Computers and Applications*, 1–5.

Suri, J., Saba, L., Dey, N., Samanta, S., Nath, S. S., Chakraborty, S., & Sanches, J. et al. (2015). Online System for Liver Disease Classification in Ultrasound. *Ultrasound in Medicine & Biology*, *41*(4Supplement), S1–S18. doi:10.1016/j.ultrasmedbio.2014.12.118

Toet, A., van Ruyven, J. J., & Valeton, J. M. (1989). Merging thermal and visual images by a contrast pyramid. *Optical Engineering (Redondo Beach, Calif.)*, *28*(7), 789–792. doi:10.1117/12.7977034

Vetterli, M., & Herley, C. (1992). Wavelets and filter banks: Theory and design. *IEEE Transactions on Signal Processing*, *40*(9), 2207–2232. doi:10.1109/78.157221

Yang, Y., Huang, S. Y., Gao, F. J., & Qian, Z. (2014). Multi-focus image fusion using an effective discrete wavelet transform based algorithm. *Measurement Science Review*, *14*(2), 102–108. doi:10.2478/msr-2014-0014

Yang, Y., Park, D. S., Huang, S., & Rao, N. (2010). Medical Image Fusion via an effective wavelet based Approach. *EURASIP Journal on Advances in Signal Processing*, 579341.

Zou, H., & Tewfik, A. H. (1992). Discrete Orthogonal M-Band Wavelet Decompositions.*IEEE International Conference on Acoustic, Speech, and Signal Processing (ICASSP)*, *4,*605-608.

This work was previously published in the International Journal of Rough Sets and Data Analysis (IJRSDA), 2(2); edited by Nilanjan Dey, pages 1-23, copyright year 2015 by IGI Publishing (an imprint of IGI Global).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 19
Analysis of Valuable Techniques and Algorithms Used in Automated Skin Lesion Recognition Systems

Uzma Jamil
Bahria University Islamabad, Pakistan & Government College University, Pakistan

Shehzad Khalid
Bahria University Islamabad, Pakistan

ABSTRACT

Application of computational intelligence techniques helps physicians as well as dermatologists in faster data process to give better and more reliable diagnoses. The whole system is categorized as: Pre-processing the lesion image to enhance its readability, Segmentation of the Lesion from skin, Feature extraction, selection, and finally the identification of dermoscopic images. Pros and cons of various methods are focused to provide a help for the researchers starting work in automated lesion detection system. Numerous computerized diagnostic systems have been reported in which different border detection, feature extraction, selection, and classification algorithms are used. The aim of this review is to summarize and compare advanced dermoscopic algorithms used for the classification of skin lesions and discuss important issues affecting the success of classification. This paper will be a guide that represents a comprehensive guideline for selecting suitable algorithms needed for different steps of automatic diagnostic procedure for ensuring timely diagnosis of skin cancer.

INTRODUCTION

In skin cancer, the ratio of malignant melanoma and non-melanoma patients is increasing day by day, in particular among people with white skin. There are many different ways of diagnoses for skin cancer which makes it difficult even for expert dermatologists to give accurate diagnosis physically. Automated lesion detection systems is the main idea used in the finding of pigmented skin lesions. By reason of

DOI: 10.4018/978-1-5225-0549-5.ch019

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 1.Skin cancer images (Ph2 Data Set, 2013)



the complication and difficulty of human understanding, automated analysis of dermoscopy images has turn into an important study and research area. An automated skin cancer detection system follows step by step procedure to detect lesion. Figure 1 is showing different types of malignant skin cancers that can lead to patient death.

To help diagnosing pigmented skin lesions, high performance computer aided investigative systems help the physicians to avoid misdiagnosis. The common approaches to skin lesion early detection include different steps of Preprocessing and Segmentation (Siegel, Naishadham, and Jemal, 2012).

The output of each step is the input of next step. Preprocessing as the first stage of computer aided cancer diagnostics has seriously effects on misleading the results (Society & A.C, 2012). The success of such systems critically depends on pre-processing (C.W.O., 2010). Preprocessing step can be divided into image enhancement, image restoration and artifact removal. Each stage includes different techniques which will be discussed in this paper. In Figure 2, a topology of automated skin lesion detection system is presented that describes the summarized information regarding first stage pre-processing, second stage segmentation, third stage feature extraction and selection and the final and most important stage classification and the details of different researchers worked on it.

This article is ordered in the following pattern. In background, previous information regarding different phases of system is explored to highlight its significance. Next stages of automated system are presented in detail. In last conclusion is described to precisely to comment the entire material.

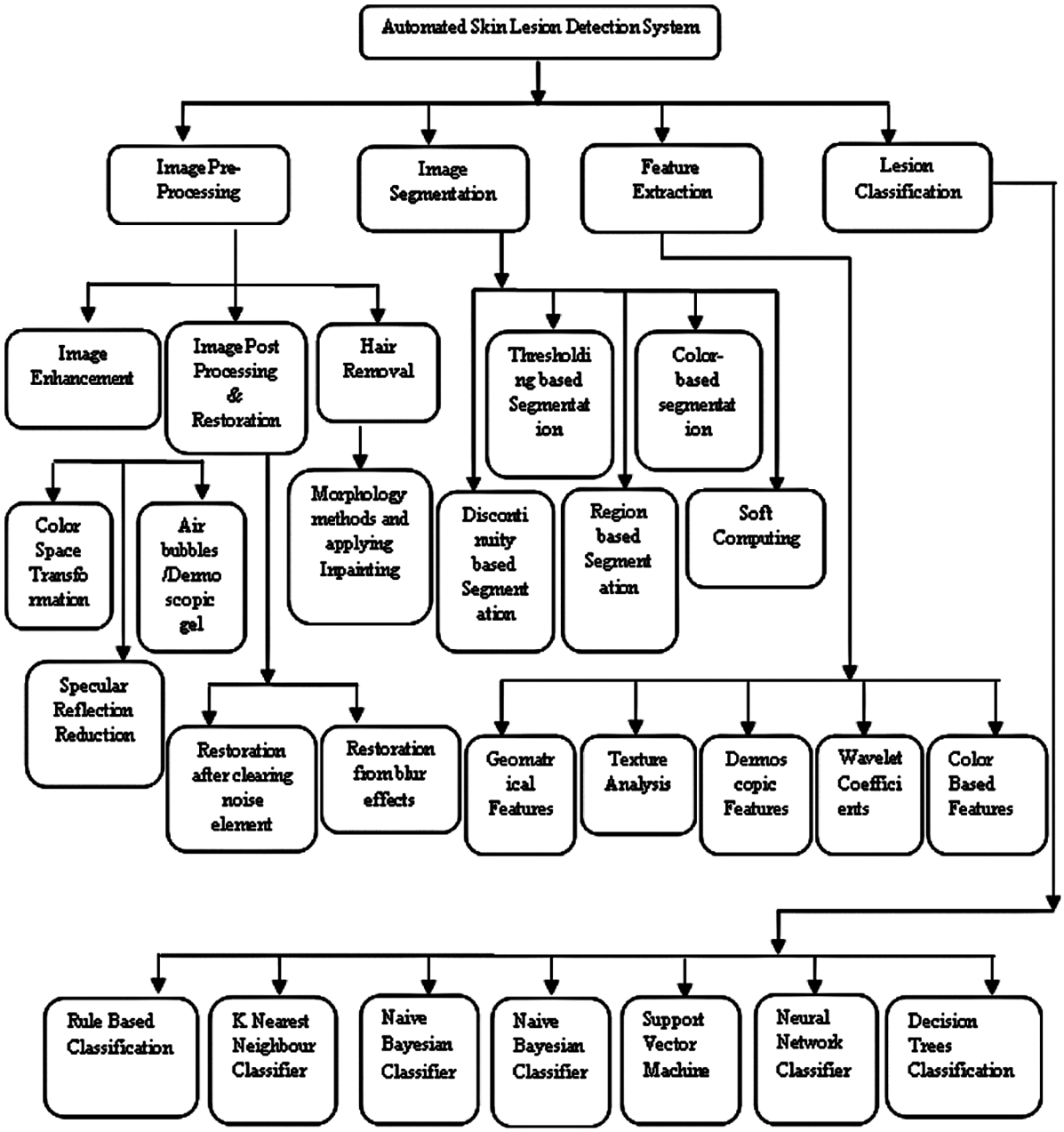
The main objective of the article is to combine the research being done related to all the steps needed for developing an automatic diagnostic system for skin cancer detection and classification. Second, it presents knowledge that help the researchers judge the importance of high level feature extraction and proper feature selection methods which needs more effort for making correct diagnosis of dangerous cancer. Third, it proposed a topology of the techniques and algorithms that are used by the researchers relevant to entire system phases to help new researchers to start with good details of techniques at one place.

Background

The frequency of skin cancer incidence is growing day by day. Estimated 1,665,540 new cases of malignant melanoma were diagnosed in USA in 2014. About 585,720 Americans are expected to die of cancer, approximately 1,600 people are dying per day (Society & A.c., 2012). Australia has one of the highest rates of skin cancer in the world. Over 1,890 Australians are dying from skin cancer every year

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 2. Topology of expository literature on automated lesion system techniques



(C.W.O, 2010). The most dangerous feature of melanoma is that it can widen broadly over the body via and blood vessels. So, early finding of skin cancer is an important factor for the prediction of the disease.

Common medical practice of skin cancer diagnosis is a manual examination by the dermatologist. Clinical diagnostic accuracy is a bit disappointing (Lindelöf & Hedblad, 1994, Morton & Machie, 1998). However, dermoscopy (Arganziano & Soyer, 2001) is a non persistent investigative system that links clinical dermatology by enabling the idea of morphological features which are not visible by analyzing manually.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Dermoscopy enables better diagnosis as compared to unaided eye (Vestergaard, Macaskill, Holt & Menzies, 2008) with an improvement in diagnostic sensitivity of 10–30. Application of computational intelligence methods helps physicians as well as dermatologists in faster data process to give better and more reliable diagnoses. After some successful experiments on automatic diagnostic systems for skin cancer diagnosis, utility of computer vision and computerized analysis is getting more important every year. Numerous computerized diagnostic systems have been reported in the literature where different border detection, feature extraction, selection, and classification algorithms are used. Some researchers reviewed and tried to critically examine image analysis techniques for diagnosis of skin cancer and compared diagnostic accuracy of experts dermoscopists with artificial intelligence and computer aided diagnosis. More research is desirable to classify and decrease uncertainties in the automatic decision support systems to improve diagnosis accuracy.

A complete state-of-the-art analysis of automatic diagnostic model for skin lesions is not available. Development of continuous new classification algorithms for dermoscopic image analysis in current years needs such an analysis. This article describes a standard automatic lesion detection system which is based on dermoscopic image examination of cancer. The aim of this review is to summarize and compare advanced dermoscopic algorithms used for the classification of skin lesions and discuss important issues affecting the success of classification. A brief and comprehensive review of feature extraction and selection algorithms that are so far being used for extracting various features of malignant melanoma is also provided. Analysis of various papers is performed with respect to several criteria, such as lesion segmentation, feature extraction, size of data sets, classification techniques, and performance measures used in reporting the diagnosis results. This paper will provide a framework that represents a comprehensive guideline for selecting suitable algorithms needed for different steps of automatic diagnostic procedure for ensuring timely diagnosis of skin cancer. The article is ordered as follows. the idea of a general computer aided diagnosis system is provided. A comprehensive review of the available literature regarding each stage is presented. The different classification algorithms are explained. Performance evaluation measures and model validation details are presented for analyzing various algorithms/models and finally concluding comments are provided.

Automated Skin Lesion Recognition System

Computer aided decision support tools are important in medical imaging for diagnosis and evaluation. By reason of the complication and difficulty of human understanding, automated analysis of dermoscopy images has turn into an important study and research area.

An automated skin cancer detection system follows step by step procedure to detect lesion. Preprocessing is the first stage, which is primarily concerned to image enhancement, artifacts removal and after that performing possible refinements, the last step of this phase is image restoration. The next phase is identifying and separating the lesion part from the skin in the provided image. Segmentation is an important phase due to accuracy rate that measures success and failure rate of an automated system. The proposed topology is presented in detail in Figure 2. Inputs to the automated lesion detection system are digital dermoscopic images. When the cancer part is segmented from the image, Morphological features can be extracted and selected for lesion recognition.

Recognition of malignant melanoma images demands very fast algorithms. A detailed research is necessary to make the best choice and to set the benchmarks for diagnostic system development and

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

validation. The following section spotlights on the explanation of the main steps that are concerned in skin cancer diagnosis.

A detailed research is necessary to make the best choice and to set the benchmarks for diagnostic system development and validation. The following section spotlights the description of the major steps and point that are concerned in skin cancer diagnosis.

Preprocessing

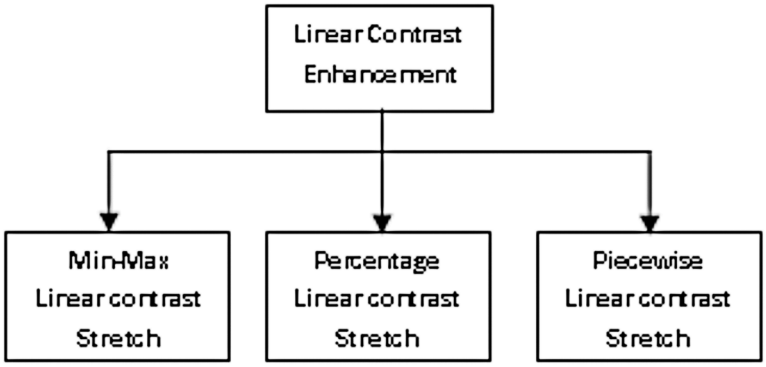
Image pre-processing is an essential step of detection in order to remove noises and enhance the quality of original image. It required to be applied to limit the search of abnormalities in the background influence on the result (Agaian, Lentz, & Grigoryan, 2000). The main purpose of this step is to improve the quality of melanoma image by removing unrelated and surplus parts in the back ground of image for further processing. Good selection of preprocessing techniques can greatly improve the accuracy of the system (TakLau, Jumaily, 2009). The objective of the preprocessing stage can be achieved through three process stages of image enhancement, image restoration and hair removal. Here, the article explains the techniques clearly for researchers who involves in preprocessing stages of automatic detections. Figure 2 is depicting the whole system stages and various techniques and algorithms that are used in lesion detection systems by different researchers.

- **Image Enhancement:** Image Enhancement is a crucial procedure to improve the visual appearance of the image; it is defined as provider of the "better" transform representation for further automated steps of detection. The image enhancement can be categorized in three categories:

    1. **Image Scaling:** Image scaling techniques are applied due to the lack of same and standard size of images. Since the skin cancer images may be gathered from different sources and sizes, the first step is to resize the images to have the fixed width pixels but variable size of height.
    2. **Color Space Transformation:** Since color information plays an inevitable role in skin cancer detection systems, researchers try to extract the more corresponding color of images for further processing. Generally, the common color spaces include RGB, HSV, HSI, CIE LAB and CIE-XYZ. RGB is a color space which comprise the red, green, and blue phantom wavelength. The most often presentation of colors in image processing is RGB. Since RGB color space has some limitation in high level processing, so other color space representations are there as an alternative. LAB is one of the useful color models which represent every color through three components of luminance, red/green and blue/yellow, it could be beneficial to transform the RGB to LAB using XYZ as an intermediate color space. The luminance would present the grayscale skin image.
    3. **Contrast Enhancement:** Contrast enhancement is beneficial step to improve the perception for further processing; it can sharpen the image border and improve the accuracy by accentuating the brightness difference between background and f foreground. Contrast enhancement plays a vital role in increasing the quality of an image. The widely practiced methods are classified into Linear contrast enhancement and Non-Linear contrast enhancement techniques. Linear contrast enhancement mostly used in remotely sensed images. This type of contrast enhancement refers to contrast stretching techniques.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 3. Linear contrast enhancement techniques

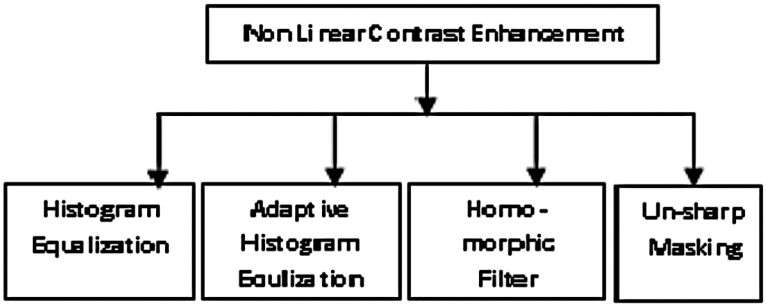


The image can be transformed to higher contrast by remapping or stretching the gray-level values so that histogram spread over the full range. Figure 3 shows the expand classification of Linear contrast Enhancement methods.

Non- Linear contrast enhancement techniques: A type of contrast improvement that mostly deals with histogram equalizations and algorithms. The most imperfection of such techniques is losing the correct brightness of an object due to the multiple values of output image against each value in an input image. In medical purposes, non-Linear contrast enhancement techniques are commonly used. The different methods of non-Linear contrast enhancement are in Figure 4.

- **Image Post Processing & Restoration:** Image Restoration is defined as the procedure to recover the degraded image from a blurred and noisy one.It can restore the degraded images in different ways. The image degradation can happen by various defects such as imperfection of imaging system, bad focusing, motion and etc. which make an image usually noisy or blur. Since the corrupted images lead to fault detection, hence, it is essential to know about noises present in an image to select the most appropriate de-noising algorithm. The image noises can be divided into four groups of Gaussian, Salt and Pepper, Poisson and Speckle.
- **Removing Thick Hairs Artifacts:** Thick hairs in automated analysis of small skin lesions are considered as a common impediment which are able to mislead the segmentation process. To remove the thick hairs in skin cancer images, researchers applied other methods such as mathemati-

Figure 4. Non-linear contrast enhancement techniques



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

cal morphology methods, curvilinear structure detection, an in-painting based method approach, automated software called DullRazor and Top Hat transform combined with a bi-cubic interpolation approach. The hair-free images are acquired using the operations. At the end of preprocessing step of skin cancer detection system, the resulting images are distinguishable from those initial images and almost are ready to feed the segmentation stage.

Segmentation

Segmentation means for dividing an image into disjoint regions which are homogeneous with respect to a property chosen as the luminance, color and texture. The aim of segmentation is to make things easier or change the representation of an image into something that is more significant and easier to examine. Some researchers disagree that manual border detection is better than the computer defined boundaries to separate the problems of feature extraction problems of computerized edge detection. However, for the development of automated diagnostic system for detection of skin lesion, it is very important to develop good segmentation algorithm. As segmentation is a critical first step in the analysis of images of the lesion, it has become one of the important research areas and many segmentation algorithms and techniques are available in the literature. However, to develop automated diagnostic system for skin lesion detection, it is very important to develop automatic segmentation algorithms. As segmentation is a crucial early step in the analysis of lesion images, it has become one of the important areas of research and many algorithms and segmentation techniques are available in the literature. We have briefly provided an overview of various segmentation algorithms being used for dermoscopic image analysis.

- **Thresholding:** Determining threshold and then the pixels are divided into groups based on that criterion. It include bi-level and multi thresholding. Thresholding method includes:
  1. Histogram
  2. Adaptive thresholding
- **Color-based segmentation algorithms:** Segmentation based on color discrimination. Include principle component transform/ spherical coordinate transform
- **Discontinuity-based segmentation:** Detection of lesion edges using active contours / radial search techniques / zero crossing of Laplacian of Gaussian (LoG). It covers:
  1. Active contours
  2. Radial search
  3. LoG
- **Region-based segmentation:** Splitting the image into smaller components then merging sub images which are adjacent and similar in some sense. It includes Statistical region merging, multi scale region growing, and morphological flooding. It is based on the following techniques:
  1. Split and merge
  2. Statistical Region Merging
  3. Multi-scale
  4. Morphological flooding
  5. Soft computing

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Soft computing techniques do the categorization of pixels using soft computing techniques. It includes:

1. Fuzzy logic
2. Neural Network
3. Optimization algorithms

Several comparative studies are also present in the literature which provides performance analysis of several segmentation algorithms. There are several issues that should be kept in mind for selecting a suitable algorithm, for example, scalar versus vector processing, automatic versus semiautomatic, and the number of parameters whose values need to be determined a priori.

Feature Extraction

It is important to identify the most effective features to extract from melanoma in situ Clark nevus and melanoma lesions, and to find the classification criteria along with most effective patterns and algorithms to differentiate the lesions. Therefore, the next stage of image analysis is to extract the important features of the image.

The reason to extract features is to decrease problem by measuring certain properties or characteristics from original data, which differentiate one from another input pattern. Feature extraction is performed by measurements on the pixels representing a segmented object that allows several features to be computed. Unluckily, the finding and extracting feature step is frequently end in error. Majority of the research is focusing on this issue, many of the features are to feed a classifying recognizer are fail to know about the true meaning of their jobs and methods to compute them. Therefore, we investigated the issue of getting lost in detail to get to the guide for future research (Kelly, Wolfe, & Simpson, 2005).

Different feature extraction methods found in the literature include model-based, filter-based and statistical methods among which the multichannel filter is more efficient and accurate. Several researchers used

- Principal Component Analysis (PCA)
- Wavelet Packet Transform (WPT)
- Co-occurrence Matrix of gray levels (GLCM)
- The Spectrum Fourier Energy
- Derived Gauss core
- Feature extraction decision boundary in order to decrease data similarity. Some of the filter banks are typically used
- Masks Laws
- Dyadic Gabor filter bank
- Wavelet transform

The main problem in the connected literature is that a many of the good research works are not elaborating their methods of feature extraction. Most important issues are concerned with the complexity in relating such information as lesion shape and color to medical elements like tissues, vessels, etc, (Celebi, et al., 2007).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Some researchers used some unique features for classification, but it is clear from research on skin cancer, a unique feature is not enough at all to detect injury and that the combination of different criteria is the key to success in early detection of malignant melanoma and other skin cancer types (She & Fish, 2003). Evolution of competing dermoscopic algorithms with variable definitions of specific attributes complicates dermoscopic diagnosis. It is necessary to identify features that are the most reproducible and diagnostically significant and formulate them into a single algorithm (Ruiz, Berenguer, Soriano, & Sánchez, 2011).

Feature Selection

Features selection is a critical step for successfully distinguishing between malignant melanoma, benign, and dysplastic nevi. Many potential features may be used, but it is important to choose a realistic decreased number of features while eliminating similar, unrelated, or noisy features. It is very important to make sure that there may not be loss of significant information.

When the classification is concerned, there are several possible benefits associated with feature selection:

- Reduced training and testing times
- Reduced feature extraction time and storage requirements
- Enhanced data understanding and visualization.
- Increased prediction accuracy
- Reduced classifier complexity for better generalization behavior

There are many methods available for feature selection (Guyon, Gunn, Nikravesh, & Zadeh, 2006), which include

- Sequential forward floating selection (SFFS)
- Principle component analysis
- Search strategies like sequential forward selection (SFS)
- Sequential backward selection (SBS)
- Fisher score ranking
- Plus-l-take-away-r (PTA (l, r))
- Sequential backward floating selection (SBFS))

All techniques need to include and exclude of progressive functions to / from the subset of concept, but these vary in their method of implementation. In fact, the floating methods are known as more smart algorithms but they are still suboptimal and there is no guarantee that they give better results. In addition to these, some of the filter-based methods include

- Mutual Information-based Feature Selection (MIFS)
- Correlation-based Feature Selection (CFS)

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Filtering methods are very fast and can estimate several alternatives in an optimization framework. It is possible to use clinical criteria or statistical methods to reduce the number of candidates variables, thus reducing the risk of a too optimistic model (Harrell & Lee, 1984).

A particular problem seen common in the relevant research is that there is very little number of studies that are presenting the detail of their feature selection criteria. Normally we do not find details of feature selection procedures that are used for choosing the appropriate features for skin cancer diagnosis.

- (Handels et al., 1999) described feature selection as an optimization problem and compared several approaches including heuristic strategies, greedy and genetic algorithms.
- (Zagrouba & Barhoumi, 2005) proposed an accelerated system for melanoma detection which is based on the function selection subset. The number of features selected by the feature selection algorithm is an important parameter. Sometimes a less number of features are not likely to discriminate between good classes. On the other hand, more features could lead to over fitting.
- (Green et al., 1994) showed by calculating correlation coefficients. Their study reveals that the size of the lesion is the most important feature for their system working.
- (Roß et al., 1995) did a feature selection by the appliance of the sequential forward selection algorithm. They attain an incredible decrease to five features starting with 87 features.
- (Celebi et al., 2007) showed by using feature selection algorithm that AUC peaks can be obtained with the use of 18 features and addition of more features does not add much to the classifier performance.
- Ganster et al., 2001) showed that the best selection performances were with subset size of between 10 to 15 and performance degrades with subsets size of more than 20 features. Moreover, inspecting the individual sensitivities of the malignant class of different sizes of subsets, it turns out that acceptable performance is achieved with subsets of more than 20 features.
- (Rohrer et al., 1998) presented a study specifically based on the selection of melanoma recognition function and showed a physically powerful boost in performance for small subsets, followed by a small addition to medium sized subsets. Large subsets cause a decrease in the recognition rate.
- (Ruiz et al., 2011) also confirmed this thing in the evaluation done using SBFS and SFFS and showed that minimum error rate was observed using subset of 6 features and a significant increase in classification error rate is observed by using a subset of more than 20 features.

The purpose of the selection function is to find the best number of features to get the top achievable performance in the classification. So it is very important that long selection techniques should be evaluated for a performance estimation on some standard classifier using ten cross-validation that is, repeating the selection of ten times with some features different data for all techniques.

Classification

Classification is a task to categorize the test sample using some already available trained data and with the help of using supervised (well known & well defined) algorithms to categorizing and finding the class of the unknown data value. Different classifying techniques are used for this purpose. In the next sections, this paper will focus and provide critical view about the following techniques: K-Nearest Neighbor, Naïve Bayesian Algorithm, Support Vector Machine, Neural networks classifier, M-mediod

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Modeling. Figure 2. is presenting the topology description of steps of an automated lesion detection system and classifiers used in the phase of classification by different researchers.

- **K Nearest Neighbor Classifier:** K-Nearest Neighbor is also known as "LAZY Learning Algorithm". The algorithm has the simplest logic. All the data set elements are the instances in some dimension space, taking the unknown sample value and getting its points in the set of space. Calculate the distance of sample value with each point in space, with any distance formula such as Euclidean distance, Minkowski, Mahalanobis city block, and chessboard. Now the next step is to sort the distances in ascending order, finding the k number of nearest points. The k should be odd for decision making. the maximum nearest neighbors will decide the class of unknown given sample. The KNN algorithm is sensitive to the garbage and irrelevant features, with slow processing time. It is quite simple and easy to implement. It also manages to handle mutually related features of any arbitrary shape. Computationally insensitive as the size of training data set increases. To improve the KNN accuracy further techniques are mixed with the algorithm such as Weighted KNN (associating weight to attributes), Backward Elimination (deletes & restores attributes) and Instance weighted KNN (Gradient Decent]. This algorithm allows the repossession and understanding of those cases that are in hand. This point is very helpful for dermatologists to directly make comparison of unknown lesions with other lesion classes.
- **Bayesian Algorithm**: Naive and simple Bayesian Algorithm is a feature based probability model. The algorithm working depend upon the probability of the instance and all the features are independent. Using the Bayesian theorem, To check the training data, compute the previous probabilities P(Ci) of each class, multiply this with likelihood i.e. P(X/Ci), then dividing it with the evidence sum P(X) of all classes. The maximum valued class later is assigned to the unknown sample data. The algorithm is fast and accurate as compare to KNN. It is not sensitive towards the garbage and irrelevant features. This is applicable for real as well as discrete data. Strong to inaccessible points and handle the missing values. It uses Gaussian standard deviation values like mean and standard deviation. The issues with Bayesian algorithm are the liberty assumption of classes, the zero probability problems and the laplacian estimator. Naive Bays Algorithm is good for any sort of features. This ability is helpful in making comparison of unknown lesion and known lesion classes. This algorithm is more conservative in estimating the performance. It avoids over-fitting and handles sampling error from different lesion datasets.
- **Support Vector Machine:** Support vector machines (SVM) are algorithms that evaluate data and identify patterns hidden in that data by applying some rules. The major concept of SVM is that it takes input and estimates, for whole data set individually, which of two possible given input groups builds the final output. For this reason it is famous as non-linear classifier. This efficient algorithm constructs a model that assigns every new data element into its group. Algorithm provide a space to mention data as points and these points are grouped so that the given classes are divided by a clear gap that is as wide as possible. Using Kernel concept, SVMs are very good for non linear classification. New data points are then mapped into that same space to predict a category based on which side of the gap they fall on. Grouping data into its particular class is a common practice that is used in medical imaging and many other fields and areas. This is the basic goal of this classifier to assign that data to its relevant group to which it belongs to. There are many different hyper planes that are used to classify the data. One logical choice to select best hyper plane

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

is that it gives maximum border or space in between classes. We select that hyper plane which is far from the nearest data point on each side.

- **Neural Network Classifier:** Complex mathematical functions are the classifiers generated by neural networks. They are rather inconceivable and dense to humans. They follow selective rule to classify data, cloudiness limits them in many applications where both accurateness and directness is the first demand, such as medical diagnosis etc. Power full data fitting or function approximation makes it vulnerable to over fitting problem. Combining several neural networks can be used to improve the performance. In many areas of life, the ANN give good results if the appropriate characteristics are used. In small picture there are many attempts to extract different types of characteristics. Since the microscopic field objects of interest have a high variability in the content, size, distribution of intensity, position, organization, form, it is difficult to extract an appropriate set of features that cover all possible situations.
- **Decision Trees:** The decision tree approach belongs to the supervised machine learning techniques. It is popular for its simplicity in constructing, efficient use in decision making, and simple representation, which is easily understood by humans. This algorithm divides repeatedly the assembly according to a criterion that maximizes the separation of the data, resulting in a tree data structure. It does this by identifying a variable and a threshold in the field of this variable can be used to divide the data set into two groups. The best choice of threshold is variable and that which minimizes the disparity of steps in the resulting groups. The most commonly used criterion is the gain information; this means that, in every group, the decrease in entropy as a result of this separation is maximized. The advantage of decision trees on most of the other methods is that they are not black box models, but can be easily expressed in rules. This makes them particularly suitable for medical applications. In many tasks classification decision tree classifiers have been preferred to other solutions (including also ANN and SVM) because they are often quick to form and generate rules apply and easy to understand. In general, Given a large set of training, decision tree classifiers generate complex decision policy that work well on the training data but do not simplify well to hidden data. In such cases, the classifier model is said to have more to adjust the training data. Another disadvantage lies in the fact that the continuous variables are implicitly discredited by the separation process, the loss of information on the way.
- **M-MEDIOD Classifier:** In this technique, the pattern is represented by a set of centers of classes uniformly disjunctive mediods. This technique has the class containing n members with M-mediods. The power of this system is its aptitude to form complex patterns without imposing any restrictions on the form of the class. It supports cluster arbitrary and very good form in the work even in the presence of outliers. When the m-mediods model for all classes have been learned, anomaly detection and classification can be done by checking the proximity of different classes using hierarchical classifier which operates in a specific hierarchy. This classifier works best for different shaped data groups and give excellent accuracy in the noise presence. High inconsistency resides in density within one group of data elements. It also identifies normality range and it also works on multi-variety settings. It can model any shape cluster and works on any feature space. This classifier is used by many medical fields such diabetic retinopathy etc and giving very good results.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 1. Feature used for making classifier comparison

Feature	Detail
Accuracy	Accuracy = No. of correct classification / Total no of Instances in the data set
Speed & Scalability	• Time to construct the model, • Time to use the model, • Efficiency in disk-resident databases
Robustness	Handling noise, missing values and irrelevant features, streaming data
Interpretability	Understanding and insight provided by the model

FEATURES USED FOR MAKING CLASSIFIER COMPARISON

The set of features used for making working comparison of these classifiers is described in Table 1 that include accuracy, Speed and scalability, Robustness and interpretability. Classification techniques are well designed algorithms that generally provide the good results of some the features but not fully accommodate the whole set of features. So it cannot be said that an algorithm lacks in some way. It is actually the problem and its environment that specifies the selection of correct classifier even use of more than one classifier is involved for accurate results in the specific problem domain. In medical imaging the ensemble classifiers are also giving very good results. These are used in dermatology, diabetic retinopathy and in many other medical imaging automated systems. More sophisticated classifiers are experienced by the researchers that specifically deal with outliers and not very much sensitive to noise and other artifacts.

In Table 2, the description of the response of classifiers against above mentioned features is presented and high-light their advantages and disadvantages when specifically used to classify injury. This is a good helper for researchers will start working on the identification of the lesion and to try to implement these classifiers. Different classification methods have their own merits. The issue of classification approach is suitable for a specific study is not easy to answer. Different classification results can be obtained by the classifier (s) selected, the differences in sample size, amount of the sample of melanoma, and the number of features used for the discrimination. Many factors, such as different sources of obtaining dermoscopic images, availability of classification software, time consumption, computational resources, and when selecting a classification method for use, the number of melanoma and benign images available for training must be acknowledge. The idea is not to rely on a unique system of decision making. Instead, all features or their subset, are used for decision making by combining their individual opinions to get a decision by consensus.

The need for data and learning techniques to make a correct diagnosis requires an appropriate choice of learning algorithms and their statistical validation. Apart from all this, to judge the performance of an automatic diagnostic model, it is important to mention who will use this model. If the automated diagnostic systems will be used by general practitioners or pharmacies and shopping centers, these systems must be used with great sensitivity and reasonably good specificity. In other words, it should recognize the greater number of melanoma at an early stage without bad nevus excision classification stage so useless benign lesions could be avoided.

If the target is the expert user, studies should be designed with the purpose to help clinicians in recognizing between benign lesions, dysplastic nevi, and malignant cancers of the skin. An increase in

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 2. Response of classifiers against features of comparison

	Accuracy	Speed & Scalability	Robustness	Interpretability
Bayesian Classifier	Exhibit high accuracy when applied to large databases.	Bayesian classifiers have also exhibited high speed when applied to large databases.	Robust to isolated noise. Handles missing values ignoring the instance during probability estimate calculation.	Simple and Naive
K Nearest Neighbor Classifier	Decreased accuracy in condensed data values and increased accuracy as data values are decreased.	The time to classify the instance is increased as the number of total instances increased.	Insensitive to noise and irrelevant transformations	Easy to understand Complexity is exponential to the no of data values.
M-Mediod Classifier	Provide good rate of Accuracy in both circular and arbitrary shaped clusters Works for general and other frame works	Efficient because works on medoids	Abnormality detection is done	Works on multi-variant settings Understand and model any shape of cluster
Support Vector Machine	Provides good accuracy in both linear and non-linear classifications	Scales well to high dimensional data	Generally insensitive to noise and other anomalies	Linear complexity

specificity might be the goal for an automated system directed to expert users together with sensitivity at least equal to that achieved by the expert.

Overall, our objective is to get a classifier with the sensibility and the specificity balanced. It should be pointed out that the skill to correctly diagnose skin lesion is important characteristic that an automated system must have. The effect of malfunction system to diagnose correctly a malignant lesion may lead to the ultimate death of the patient. On the other hand, if a classifier with a high sensibility but a low specificity is chosen, it is not going to be useful as a screening techniques.

CONCLUSION

This study gives an important contribution to this research area for several reasons. First, it is a study that combines the research being done related to all the steps needed for developing an automatic diagnostic system for skin cancer detection and classification. Second, it presents knowledge that help the researchers judge the importance of high level feature extraction and proper feature selection methods which needs more effort for making correct diagnosis of melanoma. Third, it proposed a frame work that highlights the importance of developing benchmarks and standard approaches for model validation which is generally overlooked in the previously published studies. Elegant studies are wanted to determine the accurate and suitable design features and analysis measures that are likely to lead to a good model. At this time, there are no computers that can replace an experienced clinician's intuition. Nonetheless, logic dictates that with proficient training and programming, automated systems will eventually match, if not exceed, clinical diagnostic accuracy. The refinement of current approaches and development of new techniques will help in improving the ability to diagnose skin cancer and achieving our goal of significant reduction in melanoma mortality rate.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

REFERENCES

Australia, C. W. O. (Ed.). (2010). *Causes of Death*. Canberra, Australia: Australian Bureau of Statistics.

Cancer Facts & Figures 2012. (2012). American Cancer Society. Retrieved from http://www.cancer.org/acs/groups/content/epidemiologysurveilance/documents/document/acspc031941.pdf

Celebi, M. E., Kingravi, H. A., Uddin, B., Iyatomi, H., Aslandogan, Y. A., Stoecker, W. V., & Moss, R. H. (2007). A methodological approach to the classification of dermoscopy images.[PubMed]. *Computerized Medical Imaging and Graphics*, *31*(6), 362–373. doi:10.1016/j.compmedimag.2007.01.003 PMID:17387001

Dolianitis, C., Kelly, J., Wolfe, R., & Simpson, P. (2005). Comparative performance of 4 dermoscopic algorithms by nonexperts for the diagnosis of melanocytic lesions.[PubMed]. *Archives of Dermatology*, *141*(8), 1008–1014. doi:10.1001/archderm.141.8.1008 PMID:16103330

Ganster, H., Pinz, A., Röhrer, R., Wildling, E., Binder, M., & Kittler, H. (2001). Automated melanoma recognition.[PubMed]. *IEEE Transactions on Medical Imaging*, *20*(3), 233–239. doi:10.1109/42.918473 PMID:11341712

Green, A., Martin, N., Pfitzner, J., O'Rourke, M., & Knight, N. (1994). Computer image analysis in the diagnosis of melanoma.[PubMed]. *Journal of the American Academy of Dermatology*, *31*(6), 958–964. doi:10.1016/S0190-9622(94)70264-0 PMID:7962777

Guyon, I., Gunn, S., Nikravesh, M., & Zadeh, L. A. (Eds.). (2006). *Feature Extraction: Foundations and Applications*. Dordrecht, The Netherlands: Springer; doi:10.1007/978-3-540-35488-8

Handels, H., Roß, T., Kreusch, J., Wolff, H. H., & Pöppl, S. J. (1999). Feature selection for optimized skin tumor recognition using genetic algorithms.[PubMed]. *Artificial Intelligence in Medicine*, *16*(3), 283–297. doi:10.1016/S0933-3657(99)00005-6 PMID:10397305

Harrell, F. E. Jr, Lee, K. L., Califf, R. M., Pryor, D. B., & Rosati, R. A. (1984). Regression modelling strategies for improved prognostic prediction.[PubMed]. *Statistics in Medicine*, *3*(2), 143–152. doi:10.1002/sim.4780030207 PMID:6463451

Lau, H. T., & Al-Jumaily, A. (2009). Automatically Early Detection of Skin Cancer: Study Based on Nueral Netwok Classification [sic].*International Conference of Soft Computing and Pattern Recognition*, Malacca (pp. 375-380. doi:10.1109/SoCPaR.2009.80

Lindelöf, B., & Hedblad, M.-A. (1994). Accuracy in the clinical diagnosis and pattern of malignant melanoma at a dermatological clinic.[PubMed]. *The Journal of Dermatology*, *21*(7), 461–464. doi:10.1111/j.1346-8138.1994.tb01775.x PMID:8089364

Morton, C. A., & Mackie, R. M. (1998). Clinical accuracy of the diagnosis of cutaneous malignant melanoma. *The British Journal of Dermatology, 138*(2), 283–287. doi:PubMed Argenziano, G., & Soyer, H. P. (2001). Dermoscopy of pigmented skin lesions—a valuable tool for early diagnosis of melanoma. *The Lancet Oncology, 2*(7), 443–449. doi:PubMed10.1046/j.1365-2133.1998.02075.x

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Rohrer, R., Ganster, H., Pinz, A., & Binder, M. (1998). Feature selection in melanoma recognition. *Proceedings of the 14th International Conference on Pattern Recognition* (Vol. 2, pp. 1668–1670). doi:10.1109/ICPR.1998.712040

Roß, Th., Handels, H., Kreusch, J., Busche, H., Wolf, H. H., & Pöppl, S. J. (1995). Automatic classification of skin tumours with high resolution surface profiles.*Proceedings of the 6th International Conference on Computer Analysis of Images and Patterns*, LNCS (Vol. 970, pp. 368–375). Berlin, Germany: Springer. doi:10.1007/3-540-60268-2_318

Ruiz, D., Berenguer, V., Soriano, A., & Sánchez, B. (2011). A decision support system for the diagnosis of melanoma: A comparative approach. *Expert Systems with Applications*, *38*(12), 15217–15223. doi:10.1016/j.eswa.2011.05.079

She, Z., & Fish, P. (2003). Analysis of skin line pattern for lesion classification.[PubMed]. *Skin Research and Technology*, *9*(1), 73–80. doi:10.1034/j.1600-0846.2003.00370.x PMID:12535288

Siegel, R., Naishadham, D., & Jemal, A. (2012). Cancer statistics, 2012.[PubMed]. *CA: a Cancer Journal for Clinicians*, *62*(1), 10–29. doi:10.3322/caac.20138 PMID:22237781

Vestergaard, M. E., Macaskill, P., Holt, P. E., & Menzies, S. W. (2008). Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: A meta-analysis of studies performed in a clinical setting. *The British Journal of Dermatology, 159*(3), 669–676. PubMed Agaian, S. S., Lentz, K. P., & Grigoryan, A. M. (2000). A New Measure of Image Enhancement.*International Conference on Signal Processing & Communication*.

Zagrouba, E., & Barhoumi, W. (2005). An accelerated system for melanoma diagnosis based on subset feature selection. *Journal of Computing and Information Technology*, *13*(1), 69–82. doi:10.2498/cit.2005.01.06

KEY TERMS AND DEFINITIONS

Abnormal: Apart from the expected norm, or unusual for a particular patient.

Actinic: Referring to the ultraviolet UV lamps and (UV) rays from sunlight. Sunburn is considered as an actinic burn. Actinic keratosis is a skin cancer that is the result of chronic sun contact.

Biopsy: Taking a sample of tissue for examination under a microscope to check for cancer cells or other irregularity.

Cancer: Unusual growth of cells that tend to proliferate in an uncontrolled way and, in some cases, to spread.

Dermatology: The branch of medicine concerned with the diagnosis, treatment and avoidance of diseases of the skin, hair, nails, mouth and genitals.

Diagnosis: To know about the nature of a disease; identifying a disease. A conclusion or decision of the diagnosis. The identification of any problems.

Dysplastic Nevi: A different type of moles whose look is unlike from that of a frequent normal mole. Dysplastic nevi is likely to be bigger than moles, have more unequal boundaries and are often varied in color.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Lesion: A region of abnormal tissue transform. Lesions vary in harshness from harmless to severe.

Malignant Melanoma: Melanoma is most common type of skin cancer found in people with white skin, but can be in people with all skin colors. It is the most precarious skin cancer, a tumor of the melanocyte, the cell that generate pigment in the skin.

Tumor: An unusual bunch of tissue. These bunches are a classic sign of swelling, and can be benign or cancerous.

This work was previously published in the International Journal of Privacy and Health Information Management (IJPHIM), 3(2); edited by Ernesto Jimenez-Ruiz, pages 92-107, copyright year 2015 by IGI Publishing (an imprint of IGI Global).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 20
A Study on Automatic Segmentation and Classification of Skin Lesions in Dermoscopic Images

Ebtihal Abdullah Al-Mansour
Al Imam Mohammad Ibn Saud Islamic (IMSIU), Saudi Arabia

Arfan Jaffar
Al Imam Mohammad Ibn Saud Islamic (IMSIU), Saudi Arabia

ABSTRACT

Malignant Melanoma is one of the rare and the deadliest form of skin cancer if left untreated. Death rate due to this cancer is three times more than all other skin-related malignancies combined. Incidence rates of melanoma have been increasing, especially among young adults, but survival rates are high if detected early. There is a need for an automated system to assess a patient's risk of melanoma using digital dermoscopy, that is, a skin imaging technique widely used for pigmented skin lesion inspection. Although many automated and semi-automated methods are available to diagnose skin cancer but each has its own limitations and there is no final, state-of-the art technique to date which is able to be implemented in real scenario. This survey paper is based on techniques used to segment the skin cancer, analysis of their merits and demerits and their applications on advanced imaging techniques.

INTRODUCTION

Melanoma is a type of skin cancer which consists of melanocytes (pigment-containing cells in the skin). In recent statistical, the most fatal type of skin cancer caused by Melanoma. The reason is melanoma caused for around 75% of Mortality related with skin cancer. As statistical studies in the United States, It shows that 76,690 patient with melanoma and 9,480 of them will pass away of melanoma in 2013 (Public Health Agency of Canada, 2013). In addition, in Canada, 1 in 74 male and 1 in 90 female will be infected of melanoma in their life. After applying some of studies of the melanoma trends from 1992-2006, the

DOI: 10.4018/978-1-5225-0549-5.ch020

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

researchers found that for non-Hispanic white males and females, incidence rates were increasing at an annual rate of approximately 3%. For adults; ages 15-30, melanoma is the most common diagnose types of cancer (Public Health Agency of Canada, 2013; Diepgen and Mahler, 2002). If melanoma is discovered in early stage, Stage I (less than 0.76mm thick), the rate of remaining alive for 5-years is 96%. However, if the melanoma is in Stage IV, it decreases to 5%. More and more, the treatment cost of melanoma in Stage IV is 30 times the treatment cost of melanoma in Stage I. So, early melanoma examination is helpful (Jerant, Anthony, Jennifer, Sheridan, and Timothy, 2000). Dermatoscope is a helpful tool to assist dermatologist consultants for diagnosis. It is a small device that illuminates, magnifies and enhances skin lesions, helping dermatologist to view the lesion features greatly. Comparing to the naked eye, using the dermatoscope has been improved the diagnosis. With following the ABCD scale and using a dermatoscope, diagnosis of skin lesions as malignant melanoma has a reported sensitivity of 76.0-87.7% and specificity of 61.0- 77.8%. However, in the United States, the dermatologists reported only 48% using a dermatoscope. The against using of dermatoscope causes of a lack of training or interest (Howlader et al, 2013).

The proposed research work comes under the domain of information systems for decision making activities. These systems are popularly known as Decision Support System. Decision support systems have been deployed in many areas like corporate sector, education institutions, government sector, defense organizations and medical domain etc. Decision support systems for medical applications are gaining much popularity these days among researchers and encouraging them to propose effective and efficient novel solutions to many real life problems in medical domain (Jemal et al, 2011). Medical Decision Support System have been in practice since last few decades and performing a commendable job by assisting medical and clinical experts to make successful decisions to diagnose various types of diseases. Malignant Melanoma is one of the rare and the deadliest form of skin cancer if left untreated. Death rate due to this cancer is three times more than all other skin-related malignancies combined. Incidence rates of melanoma have been increasing, especially among young adults, but survival rates are high if detected early. Unfortunately, the time and costs required for dermatologists to screen all patients for melanoma are prohibitively expensive. There is a need for an automated system to assess a patient's risk of melanoma using digital dermoscopy, that is, a skin imaging technique widely used for pigmented skin lesion inspection. One challenge in implementing such a system is locating the skin lesion in the digital image. Most existing skin lesion segmentation algorithms are affected due to the presence of noise such as illumination variation like shadows, hairs, low contrast and specular reflections in digital images. It complicates the task of finding skin lesion in the presence of these abnormalities. Existing systems on skin cancer detection are also quite costly and computationally extensive (Bleyer, O'Leary, Barr and Ries, 2006). This paper will address these problems by reviewing existing researches in order to detect skin cancer.

LITERATURE REVIEW

Segmentation of Skin Cancer

For the special problem of skin lesion segmentation, mainly region-based segmentation methods are applied, and within this category the thresholding operation is most often used. Despite of existing research in decision support systems design there exists plenty of room for researchers to explore this

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

research area and propose effective and efficient decision support systems for recognition. In medical domains (Bleyer, O'Leary, Barr and Ries, 2006), the methodology for recognition phase mostly utilizes clinical attributes for image matching. These scientific attributes provides quite useful information to medical experts which assists them in great deal to diagnose the patients. CAD systems are deployed to gain access to experts and practitioners by online access to assist the medical experts to perform their tasks efficiently (Freedberget al, 1999). These systems are developed by utilizing image visual features but major problem arises in selection of appropriate features during feature extraction phase. Therefore, this problem will be addressed in our proposed research work to extract and classify the visual features effectively.

In order to segment skin lesion images automatically, a lot of algorithms have proposed. However, like state-of-art algorithms for correction illumination, a lot of proposed segmentation algorithms can be applied only to dermoscopy images, where there is better disparity between particular types of lesions and surrounding skin area (Celebi, Emre and Hitoshi, 2009). Celebi, Emre and Hitoshi (2009) compiled a new summary of the popular existing segmentation algorithms for dermascopy images. The algorithms are compared include using active contours(Erkol et al, 2005), simple thresholding and region merging(EmreCelebi, 2008)in order to drive the segmentation, a lot of algorithms use only the features derived from pixel color(Celebi, Emre and Hitoshi, 2009). This includes the blue channel from the RGB colour space, the luminance channel from the CIE 1976 L*u*v* (CIELUV) or CIE 1976 L*a*b* (CIELAB) colour spaces, or an orthogonal transform applied to the colour channels. However, it is hard to segment lesions accurately with unclear edges when depending only on colourfeatures. The designing of these algorithms are not be worked for digital images of skin lesions. In addition, segmentation algorithms based on texture have been used to dermoscopy images. (Stoecker, Williarn, Chang-Sung, and Randy, 1992) Have been used basic statistical approaches, such as the gray-level co-occurrence matrix, to analyze texture in skin images. They found that texture analysis approach could accurately and regions with a smooth texture and that texture analysis can be applied to both segmentation and classification of dermoscopy images. For lesion segmentation, they proposed textural-based algorithms using gray-level co-occurrence matrix (Argenziano et al, 2003; Dhawan, Atam and Anne, 1992), first-order region statistics (Silveira et al, 2009), and Markov random field models (Serrano, Carmenand BegoñaAcha, 2009). The new approach proposed by (Xu, 1999) was learning a sample of natural skin texture in the four edges of the image. The existence of shadows and shining areas caused by lighting variation makes the segmentation process of skin lesion images more complicated. A few different algorithms was explored by (Hance, Gregory, Scott, Randy and William, 1996) including thresholding, active contours and split-and-merge, and change them in order to be specific to lesion images. For Example, in order to account for shining areas, the thresholding algorithm has to be changed. (Cavalcanti, Pablo, Yessenia and Jacob, 2010) provided Four separate algorithms including preprocessing step which adjust the image for lighting variation before thresholding applied or level-set segmentation algorithm (Cavalcanti, Pablo, Jacob, Leandro and Diego, 2011). Thresholding is applied on one colour channels (Cavalcanti, Pablo, Yessenia and Jacob, 2010), multiple colour channels and a group of channels derived using main component analysis and other processing steps. Including the pre-processing step helps the algorithms of lesion segmentation proposed by Cavalcanti et al. to perform perfectly. (SONALI and D.K.KAMAT, 2014) combined Thresholding segmentation technique to establishing boundaries in image with Fuzzy C-Means segmentation to find final segmentation algorithm (Manjubharathi and Saraswathi, 2014) proposed algorithm based on NC ratio analysis in automatic cell segmentation. The experimental result shows that high efficiency and accuracy of segmentation process for cancer cell. (Jeniva and Santhi,

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

2015) used the concept of texture distribution based on a learned model of natural skin and lesion textures. The texture distribution metric captures the difference between pair of texture distribution. Then, based on similarity, the images are divided into large number of smaller regions. This achieves higher segmentation accuracy. (Ram ya, 2014)the segmentation approach proposed in this paper focus on identifies skin cancer in epidermis layer of skin.The nuclei regions; which located on epidermis layer;segment using the K-means clustering algorithm based on space and some color information with k value equal to 3. After that, local region recursive segmentation (LRRS) algorithm which used intensity and size of nuclei as parameter to filter the candidate nuclei regions is performed to discover the region of nuclei. Final step is applying local double ellipse descriptor (LDED) to distinguish melanocytes from keratinocytes. This approach has good performance even if the original image is complex where background and foreground both have similar appearance. (NidhalKhdhair and Abbas, 2014)The proposed approach uses Wiener filter to remove noise such as hair from original image then, used thresholding to segment the skin cancer area from the whole image. Testing this method is provided by comparing the result of segmentation of this approach with the one done by experts in medical filed and measures the distance between these two results by using HM and TDR gives high accuracy with 96.32%. (Cheng, Muhammad, NareshJha and Mrinal, 2013) This paper proposed segmentation technique of the melanocytes in the skin histopathological image. First, using mean shift and local region recursive segmentation (LRRS) algorithm to extract nuclei areas. Then, the local double ellipse descriptor (LDED) integrates the feature of melanocytes and provides parameters to identify the melanocytes. Using 30 images with different factors as sample to test this approach showing that this technique has the ability of segmentation of melanocytes with over 80% sensitivity rate and over 70% positive prediction rate. (Binamrata, Sandeep and Toran, 2014)The proposed technique showing in this paper for segmentation is based on Neuro-Fuzzy model using decision making. Segmentation is performed with some features works as parameters. This approach gives good accuracy and quality.

Feature Extraction (Review on Automatic Early Skin Cancer Detection)

Feature extraction means extract the important feature from segmented images to be parameter for classification step. This stage is very sensitive and important step for classification. Below are different methodologies for feature extraction step. Pattern Analysis: The pattern analysis approach try to identify specific patterns, which may be global (reticular, globular, cobblestone, homogeneous, starburst, parallel, multicomponent, nonspecific) or local (pigment network, dots/globules/moles, streaks, blue-whitish veil, regression structures, hypopigmentation, blotches, vascular structures) (Ilias, Charalamposand Doukas, 2009). The ABCD-ROLE: Using four dermoscopic criteria asymmetry (A), border (B), color (C) and different structures (D) for feature extraction (Di Leo, Paolillo, Sommella, Fabbrocini and Rescigno, 2010). The ELM 7-point Check list: It is sevenstandard scoring diagnosis analysis ELM criteria includes atypical pigment network, blue-whitish veil, atypical vascular pattern, irregular streaks, irregular dots/globules, irregular blotches, and regression structures (Ilias, Charalampos and Doukas, 2009; Di Leo, Paolillo, Sommella, Fabbrocini and Rescigno, 2010). Menzies Method: This methodology seek negative features such as symmetry of pattern and presence of a single color, also positive features such as blue-white veil, multiple brown dots, pseudopods, radial streaming, scar like depigmentation, peripheral black dots/globules, multiple (five to six) colors, multiple blue/gray dots and broadened network (Ilias, Charalampos and Doukas, 2009). Texture Analysis: This method try to quantify texture notation such as fine, rough and irregular and to identify, measure and utilize the difference between them.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Classification

The final phase of the skin cancer detection system is the classification of input images to two classes (melanoma and benign). There are two main approaches used to classify dermoscopic images: The first approach; like support vector machines; assigns a label of 0 or 1 to classify the input into either melanoma or benign. The second approach; like Logistic regression, artificial neural networks, *k* -nearest neighbors, and decision trees; adds probability of class membership to the first approach in order to be the classification more accurate. In this section we have reviewed some of the classification algorithm briefly.

REVIEW ON EARLY DETECTION OF MELANOMA IN SITU

Global model such as model, linear model, polynomial models, neural networks, support vector machine Di Leo, Paolillo, Sommella, Fabbrocini and Rescigno, (2010). Semiglobal model such as redial basis feed forward network, radial basis function, multivariate adaptive regression splines (MARS), decision tree(José et al., 2009). Local model such as k - nearest neighbour model (KNN) (Bjorn, Siegfried and Albrecht Zimmermann, 2010; Skowron, Andrzej, and Arkadiusz, 2004). Hybrid model such as projection based radial basis function network (PRBFN) (Cohen, Shimon and Nathan, 2002). Below formula is used to calculate the sensitivity and specificity of above mentioned classification method:

$$sensitivity = \frac{TP}{TP + FN}$$

$$specificity = \frac{TN}{TN + FP}$$

Sachin, Shuangshuang, Atamand Dhawan, (2004); Patwardhan, Sachin, Shuangshuang and Atam (2003) evaluated the performance of classification method that is based on Neural network. The sensitivity of using this method is between 78% and 81% and its specificity in between 76% and 86%. Sachin, Shuangshuang, Atam and Dhawan, (2004) proposed a performance evaluation of classification technique based on Multivariate. This method shows that the sensitivity of it is 76% and the specificity is 65%. Ganster, (2001) and Elbaum (2001) evaluated the performance of a statistical approach for classification. The sensitivity of using this method is between 87% and 100% and its specificity in between 92% and 85%. Stanley, Moss, Van Stoecker and Aggarwal, (2003) proposed a performance evaluation of a Rule based algorithm for classification, In this algorithm the sensitivity is 80% and the specificity is 85%. Celebi, (2007) present performance evaluation of SVM approach for classification. The sensitivity in this approach is 93% and the specificity is 92%. Patwardhan, Sachin, Atam and Patricia, (2003) proposed a performance evaluation of classification technique based on ADWAT. This method shows that the sensitivity of it is 93% and the specificity is 91%. Wiltgen, Gerger and Smolle, (2003) evaluated the performance of a CART technique for classification. Both sensitivity and specificity of using this method is 92%. Ganster et al., (2001) presents performance evaluation of kNN approach for classification. The sensitivity in this approach is 87% and the specificity is 92%.The performance evaluation of

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

combining Correlation-based feature sector(CFS) + Logistic Model Trees (LMT) for classification was proposed by José et al. (2009) Using this approach shows that the sensitivity is 94% and the specificity is 68%.Table 1 shows the performance evaluation of different classification methods.

CONCLUSION

Review on Automatic Early Skin Cancer Detection

In order to find a treatment for skin cancer, the detection of this disease in early stage is important. This paper reviewed some existing methods of skin cancer detection in segmentation, feature extraction and

Table 1. Classification result as retrieved from literature review

SL No.	CLASSIFICATION METHOD	PERFORMANCE EVALUATION	
		SENSITIVITY in %	SPESIFICITY in %
1.	Neural network	78(Sachin, Shuangshuang, Atam and Dhawan, 2004), 81 (Patwardhan, Sachin, Shuangshuang and Atam,2003)	76(Sachin, Shuangshuang, Atam and Dhawan, 2004), 87(Patwardhan, Sachin, Shuangshuang and Atam,2003)
2.	Multivariate	76(Sachin, Shuangshuang, Atam and Dhawan, 2004)	56(Sachin, Shuangshuang, Atam and Dhawan, 2004)
3.	Statistical	87(Ganster, 2001), 100(Elbaum,2001)	92(Ganster, 2001), 85(Elbaum,2001)
4.	Rule based	80(Stanley, Moss, Van Stoecker and Aggarwal, 2003)	85(Stanley, Moss, Van Stoecker and Aggarwal, 2003)
5.	SVM	93(Celebi, 2007)	92(Celebi, 2007)
6.	ADWAT	93(Patwardhan, Sachin, Atam and Patricia,2003)	91(Patwardhan, Sachin, Atam and Patricia,2003)
7.	CART	92(Wiltgen, Gerger and Smolle, 2003)	92(Wiltgen, Gerger and Smolle, 2003)
8.	kNN	87(Ganster et al., 2001)	92(Ganster et al., 2001)
9.	Correlation-based feature selector (CFS) + Logistic Model Trees (LMT).	94(José et al., 2009)	68(José et al., 2009)

Table 2. Literature review

Ref	Author	Title of the Paper	Purpose	Approach	Year
(Celebi, Emre and Hitoshi, 2009)	Celebi et al.	Lesion border detection in dermoscopy images	Segmentation		2009
(Stoecker, Williarn, Chang-Sung, and Randy, 1992)	Stoecker et al.	Texture in skin images: comparison of three methods to determine smoothness	Segmentation		1992
(Xu, 1999)	Xu et al.	Segmentation of skin cancer images	Segmentation		1999
(Hance, Gregory, Scott, Randy and William, 1996)	Hance et al.	Unsupervised color image segmentation: with application to skin tumor borders	Segmentation		1996
(Cavalcanti, Pablo, Yessenia and Jacob, 2010)	Cavalcanti et al.	Pigmented skin lesion segmentation on macroscopic images	Segmentation		2010
(SONALI and D.K.KAMAT, 2014)	SONALI JADHAV	SEGMENTATION BASED DETECTION OF SKIN CANCER	Segmentation		2014

continued on following page

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 2. Continued

Ref	Author	Title of the Paper	Purpose	Approach	Year
(Manjubharathi and Saraswathi, 2014)	S. ManjuBharathi	CANCER CELL SEGMENTATION AND DETECTION USING NC RATIO ANALYSIS	Segmentation		2014
(Jeniva and Santhi, 2015)	S. Jeniva, C. Santhi	An Efficient Skin Lesion Segmentation Analysis Using Statistical Texture Distinctiveness.	Segmentation		2015
(Ram ya, 2014)	Ramya et al.	Skin Histopathological Image Used LDED Based Algarithmwith Automated Detect ion of Melanocytes Skin	Segmentation		2014
(NidhalKhdhair and Abbas, 2014)	Nidhalet al.	AUT OMAT IC SEGMENTATION OF SKIN LESIONS USING HISTOGRAM T HRESHOLDING	Segmentation		2014
(Cheng, Muhammad, NareshJha and Mrinal, 2013)	Cheng Luet al	Automated Segmentat ion of the Melanocytes in Skin Histopathological Images	Segmentation		2013
(Binamrata, Sandeep and Toran, 2014)	BinamrataBaralet al	Lesion Segmentation in Dermoscopic Images Using Decision Based Neuro Fuzzy Model	Segmentation		2014
(Ilias, Charalampos and Doukas, 2009)	IliasMaglogiannis	Overview of Advanced Computer Vision Systems for Skin Lesions Characterization	Feature extraction		2009
(Di Leo, Paolillo, Sommella, Fabbrocini and Rescigno, 2010)	G. Di Leo	A software tool for the diagnosis of melanomas, Automatic Implementation of the 7-Point Check List Method	Feature extraction		2010
(Sachin, Shuangshuang, Atam and Dhawan, 2004)	Sachin V	Multi-spectral image analysis and classification of melanoma using fuzzy membership	Classification		2004
(Patwardhan, Sachin, Shuangshuang and Atam,2003)	Z.Zhang	Neural networks skin tumor diagnostic system	Classification		2003
(Elbaum,2001)	M.Elbaum	Automatic differentiation of melanoma frommelanocytic nevi with multispectral digital dermoscopy	Classification		2001
(Ganster, 2001	H.Ganster	Automated melanoma recognition	Classification		2001
(Stanley, Moss, Van Stoecker and Aggarwal, 2003)	R.J.Stanley	A fuzzybased histogram analysis technique for skin lesion discrimination in dermatology clinical images	Classification		2003
(Celebi, 2007)	M.E.Celebi.	A methodological approach to the classification of dermoscopy images	Classification		2007
(Patwardhan, Sachin, Atam and Patricia,2003)	S.V.Patwardhan.	Classification of melanoma using tree structured wavelet transforms	Classification		2003
(Wiltgen, Gerger and Smolle, 2003)	M.Wiltgen.	Tissue counter analysis of benign common nevi and malignant melanoma	Classification		2003
(Ganster et al., 2001)	Harald Ganster	Automated Melanoma Recognition‖	Classification		2001
(José et al., 2009)	José FernándezAlcón.	Automatic Imaging System With Decision Support for Inspection of Pigmented Skin Lesions and Melanoma Diagnosis‖	Classification		2009

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

classification step. In this paper, automatic early detection of skin cancer has been divided into three main steps like preprocessing, segmentation and features extraction and classification. Different types of features extraction method has been proposed by different researchers but the best features shows the accuracy 99% are texture. Similarly different types of classifiers has been used. The most promising classifier is SVM that shows good accuracy as comparative to others.

REFERENCES

Alcón, J.F., Ciuhu, C., ten Kate, W., Heinrich, A., Uzunbajakava, N., …de Haan, G. (2009). *Automatic Imaging System with Decision Support for Inspection of Pigmented Skin Lesions and Melanoma Diagnosis. IEEE journal of selected topics in signal processing, 3(1).*

Argenziano, G., Soyer, H. P., Chimenti, S., Talamini, R., Corona, R., Sera, F., & Kopf, A. W. et al. (2003). Dermoscopy of pigmented skin lesions: Results of a consensus meeting via the Internet. *Journal of the American Academy of Dermatology*, *48*(5), 679–693. doi:10.1067/mjd.2003.281 PMID:12734496

Lu, C., Mahmood, M., Jha, N., & Mandal, M. (2013). Automated Segmentation of the Melanocytes in Skin Histopathological Images. IEEE Journal of Biomedical and Health Informatics, 17(2).

Bleyer, A., O'leary, M., Barr, R., & Ries, L.A.G. (Eds.), (2006). Cancer epidemiology in older adolescents and young adults 15 to 29 years of age, including SEER incidence and survival: 1975-2000. National Cancer Institute, Bethesda, MD, USA.

Bringmann, B., Nijssen, S., & Zimmermann, A. (2010). *From Local Patterns to Classification Models. Inductive Databases and Constraint-Based Data Mining* (pp. 127–154). Springer New York. doi:10.1007/978-1-4419-7738-0_6

Cavalcanti, P. G., Scharcanski, J., Di Persia, L. E., & Milone, D. H. (2011). An ICA-based method for the segmentation of pigmented skin lesions in macroscopic images. *Proceedings of the 2011 Annual International Conference of the IEEE on Engineering in Medicine and Biology Society EMBC* (pp. 5993-5996). doi:10.1109/IEMBS.2011.6091481

Cavalcanti, P. G., Yari, Y., & Scharcanski, J. (2010). Pigmented skin lesion segmentation on macroscopic images. *Proceedings of the201025th International Conference ofImage and Vision Computing New Zealand IVCNZ* (pp. 1-7). IEEE. doi:10.1109/IVCNZ.2010.6148845

Celebi, M., Iyatomi, H., Schaefer, G., & Stoecker, W. V. (2009). Lesion border detection in dermoscopy images. *Computerized Medical Imaging and Graphics*, *33*(2), 148–153. doi:10.1016/j.compmedimag.2008.11.002 PMID:19121917

Celebi, M.E., Kingravi, H.A., Iyatomi, H., Aslandogan, Y.A., Stoecker, W.V., Moss, R.H., & Menzies, S.W. et al. (2008). Border detection in dermoscopy images using statistical region merging. *Skin Research and Technology*, *14*(3), 347–353. doi:10.1111/j.1600-0846.2008.00301.x PMID:19159382

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Celebi, M. E., Kingravi, H. A., Uddin, B., Iyatomi, H., Aslandogan, Y. A., Stoecker, W. V., & Moss, R. H. (2007, March). A methodological approach to the classification of dermoscopy images. *Computerized Medical Imaging and Graphics*, *31*(6), 362–373. doi:10.1016/j.compmedimag.2007.01.003 PMID:17387001

Cohen, S., & Intrator, N. (2002). A hybrid projection-based and radial basis function architecture: Initial values and global optimisation. *Pattern Analysis & Applications*, *5*(2), 113–120. doi:10.1007/s100440200010

Dhawan, A. P., & Sim, A. (1992). Segmentation of images of skin lesions using color and texture information of surface pigmentation. *Computerized Medical Imaging and Graphics*, *16*(3), 163–177. doi:10.1016/0895-6111(92)90071-G PMID:1623492

Di Leo, G., Paolillo, A., Sommella, P., Fabbrocini, G., & Rescigno, O. (2010). A software tool for the diagnosis of melanomas. Proceedings of the Instrumentation and Measurement Technology Conference I2MTC. IEEE.

Diepgen, T. L., & Mahler, V. (2002). The epidemiology of skin cancer. *The British Journal of Dermatology*, *146*(Suppl. 61), 1–6. doi:10.1046/j.1365-2133.146.s61.2.x PMID:11966724

Elbaum, M., Kopf, A. W., Rabinovitz, H. S., Langley, R., Kamino, H., Mihm, M. Jr, & Wang, S. et al. (2001). Automatic differentiation of melanoma frommelanocytic nevi with multispectral digital dermoscopy: A feasibility study. *Journal of the American Academy of Dermatology*, *44*(2), 207–218. doi:10.1067/mjd.2001.110395 PMID:11174377

Engasser, H. C., & Warshaw, E. M. (2010). Dermatoscopy use by US dermatologists: A cross-sectional survey. *Journal of the American Academy of Dermatology*, *63*(3), 412–419. doi:10.1016/j.jaad.2009.09.050 PMID:20619490

Erkol, B., Moss, R. H., Joe Stanley, R., Stoecker, W. V., & Hvatum, E. (2005). Automatic lesion boundary detection in dermoscopy images using gradient vector flow snakes. *Skin Research and Technology*, *11*(1), 17–26. doi:10.1111/j.1600-0846.2005.00092.x PMID:15691255

Freedberg, K. A., Geller, A. C., Miller, D. R., Lew, R. A., & Koh, H. K. (1999). Screening for malignant melanoma: A cost-effectiveness analysis. *Journal of the American Academy of Dermatology*, *41*(5), 738–745. doi:10.1016/S0190-9622(99)70010-1 PMID:10534637

Ganster, H., Pinz, P., Rohrer, R., Wildling, E., Binder, M., & Kittler, H. (2001, March). Automated melanoma recognition. *IEEE Transactions on Medical Imaging*, *20*(3), 233–239. doi:10.1109/42.918473 PMID:11341712

Hance, G. A., Umbaugh, S. E., Moss, R. H., & Stoecker, W. V. (1996). Unsupervised color image segmentation: With application to skin tumor borders. *Engineering in Medicine and Biology Magazine, 15*(1), 104–111. doi:10.1109/51.482850

Hoshyar, A.N., Al-Jumaily, A., & Sulaiman, R. (2011). Review on automatic early skin cancer detection. *Proceedings of the2011International Conference on Computer Science and Service System CSSS* (pp. 4036-4039). IEEE.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Howlader, N., Noone, A. M., Krapcho, M., Garshell, J., Neyman, N., Altekruse, S. F., & Cronin, K. A. et al. (2013). *SEER cancer statistics review, 1975-2010 (Technical Report)*.

Baral, B., Gonnade, S., & Verma, T. (2014). Lesion Segmentation in Dermoscopic Images Using Decision Based Neuro Fuzzy Model. *International Journal of Computer Science and Information Technologies*, *5*(2), 2546–2552.

Jadhav, S.R., & Kamat, D.K. (2014). Segmentation Based Detection of Skin Cancer. *Proceedings of theIRF International Conference.*

Jemal, A., Saraiya, M., Patel, P., Cherala, S. S., Barnholtz-Sloan, J., Kim, J., & Wingo, P. A. et al. (2011). Recent trends in cutaneous melanoma incidence and death rates in the United States, 1992-2006. *Journal of the American Academy of Dermatology*, *65*(5), S17–e1. doi:10.1016/j.jaad.2011.04.032 PMID:22018063

Jeniva, S., & Santhi, C. (2015). An Efficient Skin Lesion Segmentation Analysis Using Statistical Texture Distinctiveness.

Jerant, A.F., & Johnson, J.T., Sheridan, C. & Caffrey, T.J. (2000). Early detection and treatment of skin cancer. *American Family Physician*, *62*(2), 357–386. PMID:10929700

Maglogiannis, I., & *Doukas, C.N.* (2009). *Overview of Advanced Computer Vision Systems for Skin Lesions Characterization. IEEE transactions on information technology in biomedicine, 13(5).*

Manjubharathi, S., & S. Saraswathi. (2014). Cancer cell segmentation and detection using nc ratio analysis. *International Journal of Research in Engineering and Technology.*

Nidhal Khdhair El Abbadi and Abbas Hussien Miry. (2014). Automatic segmentation of skin lesions using histogram thresholding. Journal of Computer Science, 10(4), 632-639.

Patwardhan, S. V., Dhawan, A. P., & Relue, P. A. (2003). —Classification of melanoma using tree structured wavelet transforms. *Computer Methods and Programs in Biomedicine*, *72*(3), 223–239. doi:10.1016/S0169-2607(02)00147-5 PMID:14554136

Patwardhan, S. V., Dai, S., & Dhawan, A. P. (2005). Multi-spectral image analysis and classification of melanoma using fuzzy membership based partitions. Computerized Medical Imaging and Graphics, 29(4), 287-296.

Melanoma skin cancer. (2013). Public Health Agency of Canada. Retrieved from http://www.phac-aspc.gc.ca/cd-mc/cancer/melanomaskincancer-cancerpeaumelanome-eng.php

Ramya, P. (2014). Skin histopathological image used LDED based algarithm with automated detection of melanocytes skin. *Middle-East Journal of Scientific Research*, *20*(8), 986–991.

Satheesha, T. Y., Narayana, D. S., & Giriprasad, M. N. (2012). Review on Early Detection of Melanoma in Situ. *International Journal of Advanced Technology & Engineering Research*.

Sathiya, S., Binu, S., Kumar, S., & Prabin, A. (2014). A survey on recent computer-aided diagnosis of melanoma. *Proceedings of the2014International Conference on Control, Instrumentation, Communication and Computational Technologies ICCICCT* (pp. 1387-1392). IEEE.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Serrano, C., & Acha, B. (2009). Pattern analysis of dermoscopic images based on Markov random fields. *Pattern Recognition*, *42*(6), 1052–1057. doi:10.1016/j.patcog.2008.07.011

Silla, C.N., Jr., & Freitas, A.A. (2009). A Global-Model Naive Bayes Approach to the Hierarchical Prediction of Protein Functions.

Silveira, M., Nascimento, J. C., Marques, J. S., & André, R. S. (2009). Comparison of segmentation methods for melanoma diagnosis in dermoscopyimages. *IEEE Journal of* Selected Topics in Signal Processing, *3*(1), 35–45.

Skowron, A., & Wojna, A. (2004). K nearest neighbor classification with local induction of the simple value difference metric. *Rough Sets and Current Trends in Computing*. Springer Berlin Heidelberg.

Stanley, R. J., Moss, R. H., Van Stoecker, W., & Aggarwal, C. (2003). A fuzzy based histogram analysis technique for skin lesion discrimination in dermatology clinical images. *Computerized Medical Imaging and Graphics*, *27*(5), 387–396. doi:10.1016/S0895-6111(03)00030-2 PMID:12821032

Stoecker, W. V., Chiang, C.-S., & Moss, R. H. (1992). Texture in skin images: Comparison of three methods to determine smoothness. *Computerized Medical Imaging and Graphics*, *16*(3), 179–190. doi:10.1016/0895-6111(92)90072-H PMID:1623493

Tarjoman, M., Fatemizadeh, E., & Badie, K. (2013). An implementation of a CBIR system based on SVM learning scheme. *Journal of Medical Engineering & Technology*, *37*(1), 43–47. doi:10.3109/03091902.2012.742157 PMID:23276155

Vestergaard, M. E., Macaskill, P. H. P. M., Holt, P. E., & Menzies, S. W. (2008). Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: A meta-analysis of studies performed in a clinical setting. *The British Journal of Dermatology*, *159*(3), 669–676. PMID:18616769

Wiltgen, M., Gerger, A., & Smolle, J. (2003). —Tissue counter analysis of benign common nevi and malignant melanoma. *International Journal of Medical Informatics*, *69*(1), 17–28. doi:10.1016/S1386-5056(02)00049-7 PMID:12485701

Xu, L., Jackowski, M., Goshtasby, A., Roseman, D., Bines, S., Yu, C., & Huntley, A. et al. (1999). Segmentation of skin cancer images. *Image and Vision Computing*, *17*(1), 65–74. doi:10.1016/S0262-8856(98)00091-2

Zhang, Z., Moss, R. H., & Stoecker, W. V. (2003, December). Neural networks skin tumor diagnostic system. *Proc. IEEE Int. Conf. Neural Network Signal Process* (Vol. 1, pp. 191–192).

This work was previously published in the International Journal of Knowledge Society Research (IJKSR), 6(4); edited by Miltiadis D. Lytras and Basit Shahzad, pages 51-61, copyright year 2015 by IGI Publishing (an imprint of IGI Global).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Section 3

Organizational Implications and Critical Issues

This section gives a wide variety of perspectives on oncology and research pertaining to the social and behavioral impact of oncology, for example: the corporate and managerial utilization of information sharing technologies and applications. With 9 chapters, the reader is presented with an in-depth analysis of research into the integration of global oncology as well as implementation of ethical and workflow considerations for all organizations. Chapters introducing this section critically analyze and discuss trends in oncology and crucial questions are addressed and alternatives offered along with theoretical approaches discussed. Also investigating a concern within the field of oncology is research which discusses the effect of user behavior on oncology.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 21
Role of Supercomputers in Bioinformatics

Anamika Singh
Maitreyi College, India & University of Delhi, India

Rajeev Singh
Indian Council of Medical Research, India

Neha Gupta
Northeastern University, USA & Osmania University, India

ABSTRACT

Due to the involvement of effective and client-friendly components (i.e. supercomputers), rapid data analysis is being accomplished. In Bioinformatics, it is expanding many areas of research such as genomics, proteomics, metabolomics, etc. Structure-based drug design is one of the major areas of research to cure human malady. This chapter initiates a discussion on supercomputing in sequence analysis with a detailed table summarizing the software and Web-based programs used for sequence analysis. A brief talk on the supercomputing in virtual screening is given where the databases like DOCK, ZINC, EDULISS, etc. are introduced. As the chapter transitions to the next phase, the intricacies of advanced Quantitative Structure-Activity Relationship technologies like Fragment-Based 2D QSAR, Multiple-Field 3D QSAR, and Amino Acid-Based Peptide Prediction are put forth in a manner similar to the concept of abstraction. The supercomputing in docking studies is stressed where docking software for Protein-Ligand docking, Protein-Protein docking, and Multi-Protein docking are provided. The chapter ends with the applications of supercomputing in widely used microarray data analysis.

INTRODUCTION

A supercomputer is a computer with high speed and is calculation efficient. Supercomputers first came in practice during 1960s. These supercomputers are normal and are similar to other computers but have more processors making the speed high. Presently, supercomputers are replaced by parallel supercomputers in which thousands of processors were connected to a single computer (Hoffman et al., 1990, Hill

DOI: 10.4018/978-1-5225-0549-5.ch021

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

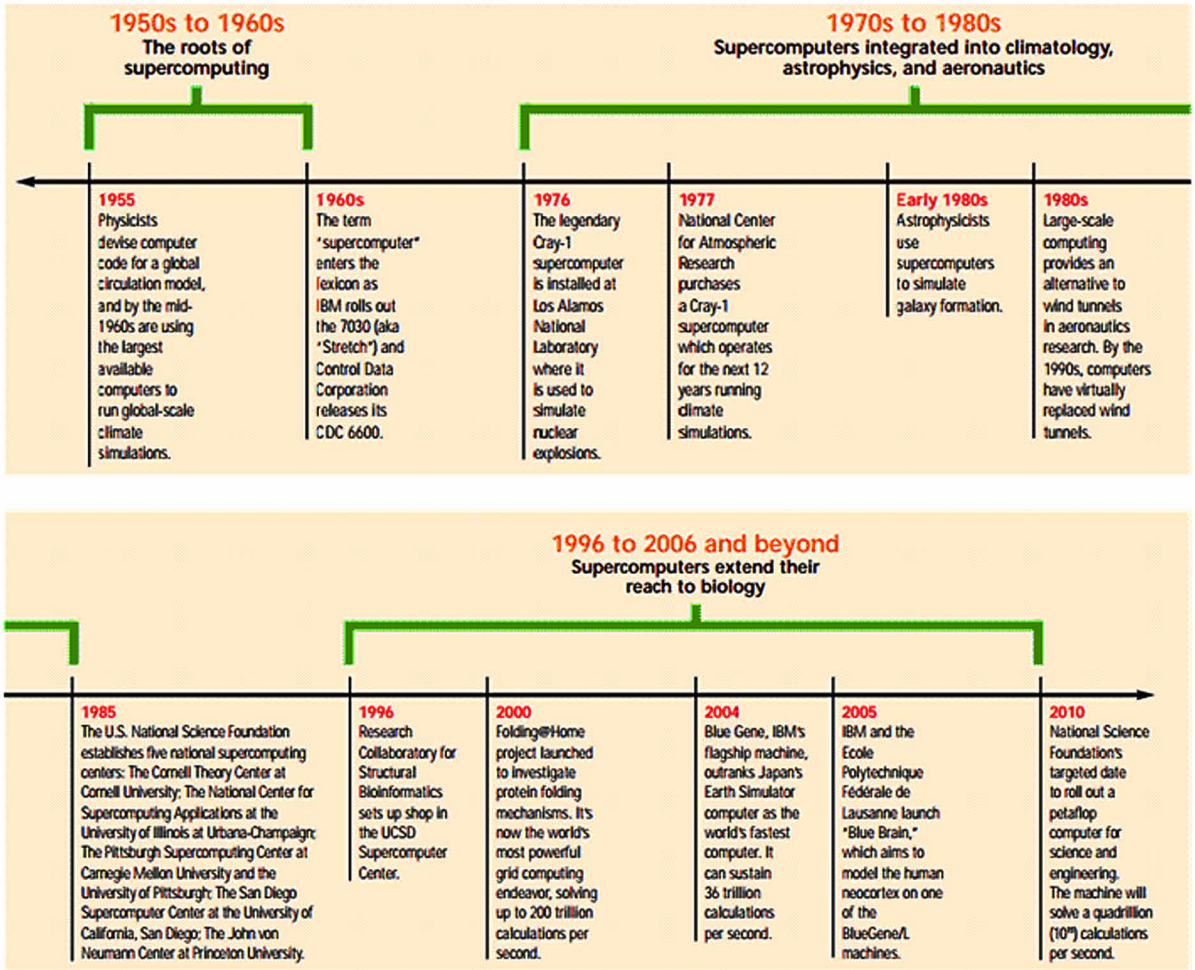
EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

et al., 1999 & Prodan et al., 2007). In this chapter we are focusing on some novel areas of biological research where supercomputing is playing a vital role. These few areas are:

SUPERCOMPUTING IN SEQUENCE ANALYSIS

Sequence analysis is actually used to explore the DNA, RNA and protein sequences in such a way that it gives all the information about the organism, source, phylogeny, function and structure, and other characteristics. Methodologies used include sequence alignment, searches against biological databases and others. Mostly it is required to search a DNA, a protein or genome database for sequence locations that are similar to that of some query sequence.

Figure 1. A brief history of supercomputing



Source: www.biomedicalcomputationreview.org

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

These databases already have billions of sequences with characteristics and this sequence information is increasing day by day. Manual searching is tough, a time consuming process and the efficiency of result is questionable. So to look for an exact match between the query string and a sub-string of the database is a very computationally demanding task. A perfect database search allows the possibility of mutations, insertion and deletions. There are novel heuristic approaches such as BLAST and FASTA (Altschul et al., 1990, Casey, 2005) which are efficient for mutations, insertions, and deletions but are not well suited for statistical purposes as they are less efficient in comparison to the dynamic programming algorithm such as Smith- Waterman (Lipman and Pearson, 1985, Pearson, W. R; Lipman, D. J. 1988). Although Smith-Waterman takes too much time for the calculations but it is still technically superior.

NCBI BLAST is the Basic Local Alignment Search Tool (BLAST) by the National Center for Biotechnology Information (Altschul et al., 1990). It is one of the most widely-used tools for sequence similarity searches. BLAST can perform comparisons between protein or DNA sequences from a sequence database where diverse sequences from different sources are present. There are different types of algorithms that were utilized for different types of search methods. There are many mathematical algorithms utilized in the analysis of sequence-sequence comparison like Genetic algorithm, Markov method, hidden Markov models, and so forth (Eddy, 1998). HMMER is also used for similarity searches of sequence databases (Edgar, 2004). Inspite of two sequences there are so many tools which can compare multiple number of sequences at a time. These are HMMER, multiple sequence alignment by CLUSTALW, Kaling etc. ("MUSCLE," http://www.drive5.com/muscle/). Sequence analysis helps molecular biology for a variety of analysis. It can compare two sequences for their similarity and identity, is helpful for the identification and analysis of active sites, interaction sites and regulatory sites. It can also identify mutations within gene and sequences. Sequence analysis also helps in genetic diversity. Here we are taking some examples of bioinformatics tools (Table 1).

SUPERCOMPUTING IN VIRTUAL SCREENING

Virtual screening is screening or filtering and identification of those molecules which have similar structures. This search is actually a database search method which contains a number of molecules with their structures and a query molecule that will match with the databases. A query may be a compound or a molecular fragment. First, the algorithm attempts to find the correct conformation and position of the ligand in the active site of the receptor and then try to quantify the quality of particular atomic arrangements by assigning a score. There are so many different methods used in protein-ligand interactions especially DOCK is a very commonly used tool for docking (Kuntz et al., 1982).

3D structure prediction of proteins and ligands gives a new insight for the prediction of accurate ligand-receptor interactions, in spite of this it can also predict the forces and bonds involved within it. 3D structure prediction is one of the most successful methods for the prediction of structures of those proteins which are not predicted by X-ray or NMR methods as these methods are having limitations. 3D structure prediction may be ab-initio, comparative methods or homology based methods and threading method.

Most successful and widely accepted method is Homology based method which is based on comparative similarity and identity based scores between query and databases. As all these methods are very popular but virtual screening is very successful only for those structures which are modeled by homology based methods (Ripphausen et al., 2010 14.).). In any case, a homology model must start from a closely related

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 1. Softwares and web based programs used for sequence analysis (Source: List of sequence alignment software, 2014)

Name	Description	Link
BLAST	local search with fast k-tuple heuristic (Basic Local Alignment Search Tool)	http://blast.ncbi.nlm.nih.gov/
FASTA	local search with fast *k*-tuple heuristic, slower but more sensitive than BLAST	http://fasta.bioch.virginia.edu/fasta_www2/fasta_list2.shtml_
HMMER	local and global search with profile Hidden Markov models, more sensitive than PSI-BLAST	http://hmmer.janelia.org/_
PSI-BLAST	position-specific iterative BLAST, local search with position-specific scoring matrices, much more sensitive than BLAST	http://blast.ncbi.nlm.nih.gov/
DNADot	Web-based dot-plot tool	www.vivo.colostate.edu/molkit/ dnadot
DOTLET	Java-based dot-plot tool	
ClustalW	Progressive alignment	www.ch.embnet.org/software/ClustalW.html
DNA Baser Sequence Assembler	Multi-alignment; Automatic batch alignment	www.DnaBaser.com_
Kalign	Progressive alignment	https://www.ebi.ac.uk/Tools/msa/ kalign/

experimental structure, so an important contributing factor in the increased utility of computational drug discovery is the rapid growth in the number of available protein structures (Joachimiak, 2009).

Another important contribution of virtual screening is generation of virtual screening databases for compounds like ZINC database at University of California San Francisco is a free database of commercially-available compounds for virtual screening. ZINC contains over 13 million purchasable compounds in ready-to-dock, 3D formats. (Irwin & Shoichet, 2005). EDinburgh University LIgand Selection System, known as EDULISS, is a relational database system for database mining.at Edinburgh University (Taylor et al., 2008). Another large database, Chemical Universe Database GDB-13 follows different methods attempting to construct the universe of synthetic compounds (Blum & Reymond, 2009).

SUPERCOMPUTING IN QSAR

Quantitative structure-activity relationship (QSAR) has been playing a major role in the field of agricultural chemistry, pharmacology, and toxicology since last few years (Hansch & Leo, 1979). With the help of QSAR, we can design new models and compare them with existing models or newly generated models to the biological databases. On the basis of similarity and dissimilarity, we can also conclude the relationship between these two.

Quantitative structure activity relationship (QSAR) modeling is an area of research which was pioneered by Hansch and Fujita (Hansch & Leo, 1995, Hansch & Selassie, 2007). The QSAR study assumes that the difference of the molecules in the structural properties experimentally measured accounts for the difference in their observed biological or chemical properties (Golender & Vorpagel 1993). With the help of QSAR, it is now possible not only to develop a model for a system but also to compare models

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

from a biological database and to draw analogies with models from a physical organic database (Hansch et al., 2001). QSAR is able to identify the relationship between a molecule and its structure and how the structure influences the activity of the molecule.

There are a few parameters like steric properties, electron distribution and hydrophobicity. In spite of this, minute analysis shows that there are some small parameters. This can affect the function of a molecule and are generally known as molecular descriptors. These descriptors are atomic descriptors and are derived from quantum chemical calculations and spectroscopy (Livingstone, 2000).

High-throughput screening method allows fast screening of large number of datasets. It separates the molecule with similar structure and function. This method is very useful as it helps to minimize the risk of comparison between different datasets of variable sources. From drug development to its mode of action, all steps can be compared easily along with drug formulation by using this method. QSAR method does not only compare the datasets but also generates the data of their analogy (Pandey & Nichols, 2011).

Latest technologies of QSAR involve three different aspects.

1. Fragment Based 2D QSAR (FB 2D-QSAR)

In this method, a drug candidate is first subjected to fragmentation into its components and then bioactivities of drug candidates are correlated with physicochemical properties of the molecular fragments through two sets of coefficients: one is for the physicochemical properties and the other for the molecular fragments (Du et al., 2008). FB-QSAR is used to predict the model of neuraminidase inhibitors for drug development against H5N1 influenza virus (Du et al., 2009). This method is also used to predict the new insights for dealing with the drug-resistant problem and designing effective adamantine-based anti-flue drugs for influenza A viruses (Wei et al., 2009).

2. Multiple Field 3D QSAR (MF 3D QSAR)

It is based on comparative molecular field analyses (CoMFA) in which two sets of coefficients are involved: one is for the potential fields and the other for the Cartesian three dimensional grid points (Du et al., 2008).

Other 3D QSAR methods include Comparative Molecular Similarity Indices Analysis (CoMSIA) that is widely used in drug design. The main advantage of this method along with CoMFA is that they can be used for heterogeneous datasets and they shortlist a 3D mapped description of favorable and unfavorable interactions according to physicochemical properties. Some of the disadvantages of these methods include 3D information and alignment of the molecular structures since there are uncertainties about different binding modes of ligands and uncertainties about the bioactive conformations (Fernandez et al., 2010).

Apart from these linear methods (CoMFA and CoMSIA), Artificial Neural Networks (ANN) and Support Vector Machines (SVMs) are able to describe nonlinear relationships which brings a more realistic picture of structure-relationship paradigm as the interactions between the ligand and its biological target must be nonlinear. Some of the latest and robust methods include Bayesian-regularized genetic neural networks (BRGNNs), Genetic Algorithm (GA) - optimized SVM (GA-SVM), Principal component-Genetic Algorithm-Artificial Neural Network (PC-GA-ANN), GA-based partial least squares (GA-PLS), etc (Fernandez et al., 2010).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

3. Amino Acid-Based Peptide Prediction (AABPP)

In this method, the biological activities of the peptides and proteins are correlated with the physicochemical properties of all or partial residues of the sequence through two sets of coefficients: one is for the physicochemical properties of amino acids and the other for the weight factors of the residues (Du et al., 2008).

QSAR and QSPR are the fields that statistically correlate chemical substance characters with biological activities (QSAR) or properties (QSPR). The chemical substance can be molecular structures, drugs, and true mixtures like nanomaterials (nanoQSAR). One of the latest tools to make QSAR more reproducible is the CHEMINF Ontology. The QSAR datasets that consists of an open XML format (QSAR-ML) is another type of QSAR study. One of the future applications would be to reformulate QSAR-ML to be based on CHEMINF (Willighagen, 2011).

AutoQSAR is a proprietory system developed at AstraZeneca in collaboration with Accelrys. Its main purpose is to automatically create, evaluate and maintain QSAR models (Cala et al.). That is to improve prediction accuracy of models by updating them with newly acquired data. AutoQSAR employs the Sun Grid Engine Platform for the calculations. Cloud computing is one area where QSAR studies can be employed and made faster to increase the speed of chemical activity predictions (Cala et al.,n.d.).

SUPERCOMPUTING IN DOCKING STUDIES

Docking is the study of close interaction between two macro molecules like proteins or it is interaction between a protein and a drug or ligand or a small peptide molecule. After interactions, these two molecules forms a stable complex and these stable complexes helps in cellular and molecular functions in cell (Lengauer & Rarey, 1996). These two molecules bind together and this interaction is based on binding affinity of two molecules. On the basis of this binding, some score was assigned to these complexes and then they were characterized on the basis of scores. So binding affinity and scoring of complexes are two important characteristics.

Docking methods are also used to predict the structure of protein-protein complex and protein-ligand complexes. So in spite of interaction studies it is also playing an important role in drug designing (Kitchen et al., 2004). Molecular docking is based on best fitted method as the ligand fits itself and it minimizes the energy of the complexes so that the complexes get stabilized itself (Jorgensen, 1991).

Another approach is induced fit method as the ligand or peptide and the receptor (protein) is flexible in nature and once ligand comes and binds with it and induces itself according to the shape of receptor molecule (Wei et al., 2000). These interactions are based on lock and key method as the receptor works as lock and the protein or the ligand molecule act as key and both gets fitted with each other and forms a stable complex (Goldman et al., 2000, Meng et al., 2004, Cerqueira et al., 2009).

Protein–Ligand Docking

Protein-ligand docking is a molecular modeling method which is used to predict the exact orientation of the ligand and protein and this orientation stabilizes the complex (Mendoza-Barrera et al., 2013). Nowadays, these interaction studies helps in pharmaceutical research as it helps in interactions as well

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

as screening of the molecules based on target's (protein) 3D structure. On the basis of these interactions, new lead molecules can be discovered and new drug molecules can be generated.

Several protein–ligand docking software applications are available, such as AutoDock or EADock. There are also web services (Molecular Docking Server, SwissDock) that calculate the site, geometry and energy of small molecules interacting with proteins. From last few decades, screening of drug and drug like molecules gave a new method of computational biology i.e. computer added drug designing (CADD) (Cerqueira et al., 2009). While dealing with receptor flexibility and degree of freedom creates problems in calculations, neglecting these problems can lead to non-reliable computational predictions (Kearsley et al., 1994).

CombiBUILD is a protein-ligand docking software developed at Sandia National Labs which is a structure-based drug design program created to aid in the design of combinatorial libraries. It screens a library of possible reactants on the computer and predicts which ones are most significant. It has been previously successful in finding nanomolar inhibitors of Cathepsin D. DockVision is a docking package created by scientists at University of Alberta using Monte Carlo, Genetic Algorithm and database screening docking algorithms. FRED (OpenEye) is an accurate and extremely fast, multiconformer docking program. FRED examines all possible poses within a protein active site, filtering for shape complementarity and optional pharmacophoric features. FlexiDock, FlexX, GLIDE, GOLD, HINT, LIGPLOT, SITUS, VEGA are other globally used efficient protein-ligand docking softwares.

Macromolecular Docking or Protein-Protein Docking

Macromolecular docking or protein-protein docking is the computational modeling of complexes. There are so many structures that can be predicted due to flexibility of the ligand and the receptor. On the basis of these flexible interactions these docked complexes were ranked and some scores were assigned to them and based on their molecular affinity best docked complex can be isolated as a final docked complex.

Docking may be rigid or flexible depending upon their motions at molecular level. Conformations of the ligand may be generated and further this model is used for the interaction studies when it will docked with another protein or receptor (Friesner et al., 2004). These ligands are docked at the cavity of the receptor where it gets a space and has fitted itself in such a way that it also minimizes the energy of the complex (Zsoldos et al., 2007).

Another approach of docking is fragment based docking as it gives a small fragment to dock with a small portion of the receptor (Wang & Pang, 2007, Klebe & Mietzner, 1994, Cerqueira et al., 2009). Flexibility in receptors always create problems in analysis as they have a bulky structure and every molecule is having their own degree of freedom (Totrov & Abagyan, 2008, Hartmann et al., 2009, Taylor et al., 2003).

There are many interesting aspects which influence the protein-protein interaction mostly like effects of amino acids like Arginine, Histidine and Lysine increases the positive charge which causes charge-charge repulsion and finally decrease of protein stability. Protein-Protein interactions mainly include electrostatic interactions, hydrogen bonds, the Vanderwalls interactions and hydrophobic interactions. Average protein-protein interface is not less polar or more hydrophobic than the surface remaining in contact with solvent. Hydrophobic forces derive protein-protein interactions and hydrogen bonds and salt bridges generally confer specificity. The Vanderwalls forces bind between the neighboring atoms if they are numerous and tightly packed they do contribute to the binding energy of association.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

It was also observed by the earlier workers that hydrogen and Vanderwall bonds are more favorable between protein molecules than the surrounding water molecules. Shape of interacting surface also influences the interactions generally interactive surface is flat in nature. There are some more quite interesting structural factors noticeable in the Protein-Protein interactions like loop regions of proteins are more involved in the interactions.

MEGADOCK 3.0 is a high-throughput and ultra-fast Protein-Protein Interaction prediction software based on rigid docking. It employs an MPI/OpenMPI technique assuming usages on massively parallel supercomputing systems. MEGADOCK displays significantly faster processing speed that leads to full usage of protein tertiary structural data for large-scale problems in systems biology. This is a new protein-protein docking engine aimed at exhaustive docking of mega-order numbers of protein pairs (Matsuzaki, 2013).

ClusPro 2.0 is a protein-protein docking software developed at Boston University. In each docking job, the receptor and ligand has to be chosen. The receptor can be either a structure present in the PDB or you can also upload your own PDB by clicking on the Upload PDB tab. The chains can also be specified for the receptor. Similarly, ligand is chosen. Depending on the queue of the jobs on the supercomputer and the size of the protein, docking will usually finish in a few hours. The results of docking can be viewed and downloaded. To analyze the results, the model is downloaded and can be viewed in PyMol (ClusPro-protein-protein docking).

Global Range Molecular Matching (GRAMM) at the University of Kansas is a Protein-Protein docking and Protein-Ligand docking software that uses an empirical approach to smooth the intermolecular energy function (Structure based drug design and molecular modeling). It requires only the atomic coordinates of the two molecules to predict the complex structure and therefore no binding site information is needed. It performs an exhaustive 6 –dimensional search through the translations and rotations of the molecules. ICM-Dock at MolSoft LLC is another Protein-Protein and Protein-Ligand docking software which is fast and accurate. It has a unique set of tools for accurate individual ligand-protein docking, peptide-protein docking and protein-protein docking including interactive graphic tools (Structure based drug design and molecular modeling).

Other popular protein-protein docking softwares include ZDOCK, HEX, PIPER, GRAMM-X, PRISM, 3D-Dock Suite, BiGGER, DOT, ESCHER NG, HADDOCK, etc.

Some Docking Softwares

Binding Affinity Prediction of Protein-Ligand (BAPPL) server basically computes the binding free energy of a non-metallo protein-ligand complex using an all atom energy based empirical scoring function. BAPPL server accepts two methods (Jayaram, 2012). One method is where the input should be an energy minimized protein-ligand complex with hydrogens added, protonation states, partial atomic charges and VanderWaals parameters assigned for each atom. The server directly computes the binding affinity of the complex. The second method is where the input is an energy minimized protein-ligand complex with hydrogens added and protonation states assigned. The net charge on the ligand should be specified. The server obtains the partial atomic charges of the ligand using the AM1-BCC procedure and GAFF force field for VanderWaals parameters (Jayaram, 2012).

ParDOCK is another automated server for Protein Ligand Docking at IIT Delhi. It is an all-atom energy based Monte Carlo, protein ligand docking implemented in a automated parallel processing mode. The

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

structural input data for the ParDOCK are optimized reference complex (protein-ligand) and a example ligand to be docked (Jayaram, 2013).

DnaDOCK is another bioinformatics software that does DNA ligand docking. It is again a Monte Carlo based docking implemented in a parallel processing mode and predicts the binding mode of the ligand in the minor groove of DNA (Jayaram, 2011). Superimpose is another software that fits two molecular structures drug and protein and calculates the Root Mean Square Deviation between them giving an idea about their structural similarity. RMSD is the square root of the average of the squared distance between each mapped pair. It is normally used in 3D geometry of molecules to measure distance between a given set of points or atoms. The resultant superimposed structures can be downloaded and viewed in a visualization software like ViewerLite, Rasmol, etc (Superimpose, 2010).

Multi Protein Docking

Multi Protein Docking involves the docking of many proteins of important function to a single drug candidate. There are various advantages of this docking method, some of which are

1. Determines toxicity and side effects.
2. Predicts failures earlier in the process.
3. Increases overall success rate (Ellingson, 2012).

Multi protein docking also has its own limitations. For instance, using Autodock4 software, there are separate MPI jobs for each receptor. Also, binary grid files are generated for each receptor. Hence, using supercomputing abilities in future, a tool is needed that allows an increase in the number of receptors used in a screening with a minimal increase in the amount of I/O per docking task (Ellingson, 2012).

In comparison to Autodock4, Autodock Vina is found to be potential multi-protein docking software. It calculates grid maps efficiently during docking and does not store them on disk. It is already multi-threaded, has maximum number of rotatable bonds, etc (Ellingson, 2012).

In the future, it is believed that all possible drugs can be docked into the entire set of proteins with an individual's genetic variations in order to develop personalized medicine (Ellingson, 2013). The program VinaMPI has been developed to enable large virtual drug screens using a large number of cores. It is based on the multi-threaded virtual docking program Autodock Vina. VinaMPI efficiently handles multiple proteins in a ligand screen, allowing for high-throughput inverse docking that can improve the efficiency of drug discovery pipeline. VinaMPI successfully ran on 84,672 cores with a decrease in job completion time and increasing core count (Ellingson, 2013).

Other multi-purpose docking softwares include InsightII, AMBER, GRASP, Procheck, GCG program, HOMDOCK, ICM, ParaDockS, Surflex-Dock, LigandFit, etc.

Supercomputing in Microarray Data Analysis

Microarray is a technique in which DNA probes that are arrayed on a solid support (Silicon thin film) are used for assay. This analysis is based on the hybridization ability of the DNA with the probe as the probe is designed, in such a manner that it must have complementary sequences (DNA) (Marmur & Doty, 1961). Microarray is a chip of very small size having ninety six or more tiny wells and each well

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

has thousands of DNA probes or oligonucleotides arranged in a grid of chip ((Sundberg et al., 2001), (Afshari & Perspective, 2002)).

Different types of genes are immobilized and are fixed at specific locations on chip and thus a single chip can give the information about thousands of genes simultaneously by hybridization method. cDNA microarrays and oligonucleotide arrays are the two types of microarray data analysis ((Ponder, 2001), (Rowley,1973)).

1. cDNA arrays are generated by putting a double stranded cDNA on a solid support (glass or nylon). For this robotics arms are used for the minimization of errors.
2. Oligonucleotide arrays are made by synthesizing specific oligonucleotides in a specific alignment on a solid surface. For this photolithography technique is used. (Gray et al., 2000).

The labeled cDNAs are now exposed and allowed to hybridize with the probe DNA. This hybridization is similar to Southern blotting method. After that slide is washed properly which removes nonspecific hybridization, it is read in a laser scanner that can differentiate between Cy3- and Cy5-signals, collecting fluorescence intensities to produce a separate 16-bit TIFF image for each channel ((Venter & Adams, 2001), (Young, 2000)).

Estimation of results is done by measuring the intensity of fluorescence, which corresponds to the amount of gene expressed in the sample (Pandey et al., 2003 & Labana et al., 2005). The three major steps of a microarray technology are preparation of microarray, preparation of labeled probes and hybridization and finally, scanning, imaging and data analysis ((Labana et al., 2005), (Esteve-Nunez et al., 2001)).

The scanning and data analysis all depends on the efficiency of supercomputers.

The following steps are important for the accurate data quantification.

- The multiple levels of replication in experimental design (Experimental design)
- The number of platforms and independent groups and data format (Standardization)
- The treatment of the data (Statistical analysis with the help of supercomputers)
- Accuracy and precision (Relation between probe and gene with the help of supercomputers)
- The sheer volume of data and the ability to share it (Data warehousing with the help of supercomputers)

Applications of DNA Microarray Technology

DNA microarray technology has been used to study many bacterial species which include *Escherichia coli* (Richmond et al., 1999, Tao et al., 1999). *Mycobacterium tuberculosis* (Wilson et al., 1999; Behr et al., 1999), *Streptococcus pneumonia* (De Saizieu et al., 2000 & Hakenbeck et al., 2001) and *Bacillus subtilis* ((Ye et al., 2000), (Yoshida, 2001)). With DNA microarray entire microbial genome can be easily represented in a single array and it is feasible to perform genome-wide analysis (DeRisi et al., 1997). Microarray technique is used in medicine development by providing microarray data of a patient which could be used for identifying diseases (Amandeep Singh et al., 2013). DNA microarray technology has been used for analyses of natural and anthropogenic factors in yeast and analyzed how the whole genome of yeast responds to environmental stressors such as temperature, pH, oxidation, and nutrients ((Causton et al., 2001), (Gasch et al., 2000)).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Microarray analysis has been applied to identify molecular markers of pathogen infection in salmon (Rise et al., 2004). DNA microarray has been used for studying gene expression analysis in neurological disorders. DNA microarray experiments are carried out to find genes which are differentially expressed between two or more samples of cells (Caetano et al., 2004). cDNA microarrays provide a powerful tool for studying complex phenomena of gene expression patterns in human cancer (Ghosh et al., n.d.). DNA microarrays has been used for clinical diagnosis such as histopathology and molecular pathology e.g. Microarray technique has been identified for analysis of AMACR (α-methylacyl-CoA racemase in prostate cancer compared with normal prostate. ((Shalon et al., 1996), (Schena et al., 1996)).

Scientists at the Houston Methodist Research Institute (HMRI) have been successful in understanding the inverse relationship between Alzheimer's disease and a brain cancer, Glioblastoma Multiforme (GBM) (Dutt,2014). Information about the genes were obtained by DNA microarray analysis followed by sequencing using data from The Cancer Genome Atlas at NIH (for GBM) and Alzheimer's Disease Neuroimaging Initiative (for AD).The results were then analyzed by supercomputers, which yielded a 1000-fold improved performance as compared to microarray analysis (Dutt, 2014).

Analysis of raw microarray data files, e.g. Affymetrix CEL files can be time consuming, complex and requires fundamental computational skills. Also, the analysis of microarray data normally involves many steps like data organization, quality control, normalization, differential gene expression calling, clustering, pathway analysis, etc. The development of software-based workflow to automate these procedures would improve the efficiency and serve to standardize the multiple inter and intra-dataset analyses.

Typical microarray-based gene expression analyses compare gene expression in adjacent normal and cancerous tissues. In this type of analysis, genes with strong statistical differences in expression are identified (O'Neal, 2014). However, many genes are aberrantly expressed in tumors. These "passenger" genes are differentially expressed between normal and tumor tissues but they are not "drivers" of tumorigenesis. Therefore, better computational methods that enrich the list of genes with authentic cancer-associated "driver" genes are needed (O'Neal, 2014).

Multicategory Support Vector Machines (MC-SVM) are powerful classification systems with excellent performance in a variety of problems associated with data classification. Since the process of generating models in traditional Multicategory Support Vector Machines for large datasets is very computationally intensive, there is a need to improve this performance using advanced High Performance Computing techniques (Zhang et al., 2006). PMC-SVM is developed for classifying large microarray datasets. It is implemented in MPI and C++ based on the serial implementation of SMO-SVM in libSVM. The experimental results show that the high performance computing techniques and parallel implementation can achieve a great speed-up without losing accuracy (Zhang et al., 2006).

The completed human genome consists of ~ 3000 mega-bases and ~35000 genes distributed across 23 chromosomes. The mouse genome is similar in size and gene number except with 3 fewer chromosomes (Bertone, 2014). This genetic data is a good standard in comparative genome analysis to annotate gene information of other organisms. However, huge computation is involved in sequence similarity comparisons. Field Programmable Gate Arrays (FPGA) are integrated circuits designed to perform far more quickly than conventional CPUs. In a way, FPGAs are similar to the application-specific integrated circuits (ASICs). TimeLogic has already developed DeCypher system based on the FPGA technology with enhanced Biocomputing algorithms including BLAST for genome level analyses. This is a unique DNA microarray design algorithm using the FPGA technology combined with de novo software development (Bertone, 2014).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Many existing Bioinformatics tools like Bioconductor, BASE, etc have been developed to assist in the analysis of microarray datasets. The available tools for microarray analyses are developed in Perl, Python and R programming languages (Gan et al., 2014). Apart from these tools, different platforms like Galaxy, Taverna, Bioclipse, Yabi and Kepler are available to develop Bioinformatics pipelines. Kepler is the most popular as it has a user friendly graphical interface, built-in R/python components, etc. The ability to extend workflows as well as the support of external packages such as Bioconductor makes Kepler an ideal platform to work on further analytic tools. For instance, Kepler based workflow was developed by integrating desired tools to develop a pipeline termed Meta-Analyses of online-available Affymetrix Microarray Data (MAAMD). MAAMD is the standard microarray analysis pipeline and also improves analysis efficiency. Some of the tools that are integrated include Bioconductor packages like AffyQCReport and arrayQualityMetrics, AltAnalyze (for Microarray and RNA-Seq analysis), etc (Gan et al., 2014).

There are several Microarray Analysis softwares, some of them are GoMiner, GoSurfer, GenMAPP, ArrayTrack, caGEDA, SAM, NUDGE, etc. The Microarray Data Repositories include GEO, ArrayExpress, CIBEX, Standard Microarray Database (SMD), The Gene Expression Database (GXD), Oncomine, etc (Yang, 2006). An approach to gene expression analysis must involve an up-front characterization of the structure of data. Singular Value Decomposition (SVD) and Principle Component Analysis (PCA) can be valuable tools in obtaining such characterization (Wall,n.d.). SVD and PCA are common techniques for analysis of multivariate data and gene expression data are well suited to analysis using SVD/PCA. Gene expression data are normally very noisy and SVD can detect and extract small signals from noisy data. PCA and SVD analysis methods are also useful for characterizing protein molecular dynamics trajectories and visualization of the datasets. Other techniques to visualize data are Multidimensional Scaling and Self Organizing Maps (SOM) (Wall, n.d.).

The statistical tests are normally done in order to find out the differentially expressed genes. Basically, two types of tests are performed on microarray data. They are t-test and ANOVA test. A more sophisticated approach, Bayesian Belief Networks (BBN) or Bayesian probabilistic framework is useful to analyze expression data (Valafar, 2002). In this method, competing models are designed and then Bayesian Belief Networks is used to pick the model that best fits the expression data. Friedman and co-workers have used BBNs to analyze genome-wide expression data in order to identify significant interactions between genes in a variety of metabolic and regulatory pathways (Valafar, 2002). Bayesian Neural Networks (BNN) is another technique that is used for gene expression analysis. Liang et al have used the BNNs with structural learning for exploring microarray data in gene expressions (Valafar, 2002).

CONCLUSION

Supercomputers are used for fast calculations and accurate predictions. By using these methods, we can predict so many aspects of biological molecules which are not possible experimentally due to some limitations. Computational biology, drug designing, Sequencing, genomic analysis, protein 3D structure prediction are the major areas of biological research. Neural networks and systems biology, artificial neural networks and cloud computing are the new emerging branches of science which are based on supercomputers.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

REFERENCES

Afshari, C. A. (2002). Perspective: Microarray technology, seeing more than spots. *Endocrinology*, *143*(6), 1983–1989. doi:10.1210/endo.143.6.8865 PMID:12021158

Altschul, S., Gish, W. M., Myers, W., & Lipman, E. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. doi:10.1016/S0022-2836(05)80360-2 PMID:2231712

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. doi:10.1016/S0022-2836(05)80360-2 PMID:2231712

Behr, M. A., Wilson, M. A., Gill, W. P., Salamon, H., Schoolnik, G. K., & Rane, S. et al. (1999). Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science*, *284*(5419), 1520–1523. doi:10.1126/science.284.5419.1520 PMID:10348738

Bertone, A. (2014). *FPGA assisted equine gene annotation and custom microarray (genechip)*. Retrieved June 16, 2014 from https://www.osc.edu/research/bioinformatics/projects/horse_gene

Blum, L. C., & Reymond, J. (2009). Million drug like small molecules for virtual screening in the chemical universe database GDB-13. *Journal of the American Chemical Society*, *131*(25), 8732–8733. doi:10.1021/ja902302h PMID:19505099

Caetano, A. R., Johnson, R. K., Ford, J. J., & Pomp, D. (2004). Microarray profiling for differential gene expression in ovaries and ovarian follicles of pigs selected for increased ovulation rate. *Genetics*, *168*(3), 1529–1537. doi:10.1534/genetics.104.029595 PMID:15579704

Cala, J., Hiden, H., Watson, P., & Woodman, S. (n.d.). *Cloud computing for fast prediction of chemical activity*. Retrieved June 16, 2014 from http://www.esciencecentral.co.uk/wp-content/uploads/2011/03/Cloud-Computing-for-Fast-Prediction-of-Chemical-Activity.pdf

Casey, R. M. (2005). *BLAST sequences aid in genomics and proteomics*. Business Intelligence Network.

Causton, H. C., Ren, B., Koh, S. S., Harbison, C. T., Kanin, E., & Jennings, E. G. et al. (2001). Remodeling of yeast genome expression in response to environmental changes. *Molecular Biology of the Cell*, *12*(2), 323–337. doi:10.1091/mbc.12.2.323 PMID:11179418

Cerqueira, N. M., Bras, N. F., Fernandes, P. A., & Ramos, M. J. (2009). MADAMM: A multistaged docking with an automated molecular modeling protocol. *Proteins*, *74*(1), 192–206. doi:10.1002/prot.22146 PMID:18618708

Cerqueira, N. M., Fernandes, P. A., Eriksson, L. A., & Ramos, M. J. (2009). MADAMM: A multistaged docking with an automated molecular modeling protocol. *Proteins: Structure, Function, and Bioinformatics*, *74*(1), 192–206. doi:10.1002/prot.22146 PMID:18618708

ClusPro-Protein-Protein Docking. (n.d.). Retrieved June 16, 2014 from http://cluspro.bu.edu/tut_dock.php

De, K., Ghosh, G., Datta, M., Konar, A., Bandyopadhyay, J., & Bandyopadhyay, D. et al. (2004). Analysis of differentially expressed genes in hyperthyroid-induced hypertrophied heart by CDNA microarray. *The Journal of Endocrinology*, *182*(2), 303–314. doi:10.1677/joe.0.1820303 PMID:15283691

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

de Saizieu, A., Gardes, C., Flint, N., Wagner, C., Kamber, M., & Mitchell, T. J. et al. (2000). Microarray-based identification of a novel *Streptococcus pneumoniae* regulon controlled by an autoinduced peptide. *Journal of Bacteriology*, *182*(17), 4696–4703. doi:10.1128/JB.182.17.4696-4703.2000 PMID:10940007

DeRisi, J. L., Iyer, V., & Brown, P. O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*, *278*(5338), 680–686. doi:10.1126/science.278.5338.680 PMID:9381177

Du, Q. S., Huang, R. B., & Chou, K. C. (2008). Review: Recent advances in QSAR and their applications in predicting the activities of chemical molecules, peptides and proteins for drug design. *Current Protein & Peptide Science*, *9*(3), 248–259. doi:10.2174/138920308784534005 PMID:18537680

Du, Q. S., Huang, R. B., Wei, Y. T., Pang, Z. W., Du, L. Q., & Chou, K.-C. (2009). Fragment-Based quantitative structure-activity relationship (FBQSAR) for fragment-based drug design. *Journal of Computational Chemistry*, *30*(2), 295–304. doi:10.1002/jcc.21056 PMID:18613071

Dutt, S. (2014). *HMRI researchers use systems biology, TACC supercomputers to find link between Alzheimer's and brain cancer*. Retrieved June 16, 2014 from http://bionews-tx.com/news/2014/04/29/alzheimers-and-cancer-link-found/

Eddy, S. R. (1998). Profile hidden markov models. *Bioinformatics (Oxford, England)*, *14*(9), 755–763. doi:10.1093/bioinformatics/14.9.755 PMID:9918945

Edgar, R. (2014). *Muscle*. Retrieved June 16, 2014 from http://www.drive5.com/muscle/

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. doi:10.1093/nar/gkh340 PMID:15034147

Ellingson, S. (2012). *Accelerating virtual high-throughput ligand docking*. Retrieved June 16, 2014 from http://salsahpc.indiana.edu/ECMLS2012/slides/ECMLS12_Accelerating_Virtual.pdf

Ellingson, S. (2013). Multi-receptor high-throughput virtual docking on supercomputers with VinaMPI. Paper presented at SC13 2013, Denver, Colorado.

Esteve-Nunez, A., Caballero, A., & Ramos, J. L. (2001). Biological degradation of 2, 4, 6-trinitrotoluene. *Microbiology and Molecular Biology Reviews*, *65*(3), 335–352. doi:10.1128/MMBR.65.3.335-352.2001 PMID:11527999

Fawcett, P., Eichenberger, P., & Losick, R. (2000). The transcriptional profile of early to middle sporulation in Bacillus subtilis. *Proceedings of the National Academy of Sciences of the United States of America*, *97*, 8063–8068. doi:10.1073/pnas.140209597

Fernandez, M., Caballero, J., Fernandez, L., & Sarai, A. (2010). *Genetic algorithm optimization in drug design QSAR: Bayesian-regularized genetic neural networks (BRGNN) and genetic algorithm-optimized support vectors machines (GA-SVM)*. Retrieved June 16, 2014 from http://www.researchgate.net/publication/42373086_Genetic_algorithm_optimization_in_drug_design_QSAR_Bayesian-regularized_genetic_neural_networks_(BRGNN)_and_genetic_algorithm-optimized_support_vectors_machines_(GA-SVM)

Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., & Mainz, D. T. et al. (2004). Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of Medicinal Chemistry*, *47*(7), 1739–1749. doi:10.1021/jm0306430 PMID:15027865

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Gan, Z., Stowe, J., Altintas, I., McCulloch, A., & Zambon, A. (2014). Using kepler for tool integration in microarray analysis workflows. In *Proceedings of 14th International Conference on Computational Science* (pp. 2162-2167). Academic Press. doi:10.1016/j.procs.2014.05.201

Gasch, A. P., Spellman, P. T., Kao, C. M., Carmel-Harel, O., Eisen, M. B., & Storz, G. et al. (2000). Genomic expression programs in the response of yeast cells to environmental changes. *Molecular Biology of the Cell*, *1*(12), 4241–4257. doi:10.1091/mbc.11.12.4241 PMID:11102521

Goldman, B. B., & Wipke, W. T. (2000). QSD quadratic shape descriptors. 2. Molecular docking using quadratic shape descriptors (QSDock). *Proteins*, *38*(1), 79–94. doi:10.1002/(SICI)1097-0134(20000101)38:1<79::AID-PROT9>3.0.CO;2-U PMID:10651041

Golender, V. E., & Vorpagel, E. R. (1993). In H. Kubinyi (Ed.), *In 3D-QSAR in drug design: Theory, methods, and application* (p. 137). ESCOM Science Publishers.

Gray, J. W., & Collins, C. (2000). Genome changes and gene expression in human solid tumors. *Carcinogenesis*, *21*(3), 443–452. doi:10.1093/carcin/21.3.443 PMID:10688864

Hakenbeck, R., Balmelle, N., Weber, B., Gardes, C., Keck, W., & de Saizieu, A. (2001). Mosaic genes and mosaic chromosomes: Intra- and interspecies genomic variation of *Streptococcus pneumoniae*. *Infection and Immunity*, *69*(4), 2477–2486. doi:10.1128/IAI.69.4.2477-2486.2001 PMID:11254610

Hansch, C., Kurup, A., Garg, R., & Gao, H. (2001). Chem-bioinformatics and QSAR: A review of QSAR lacking positive hydrophobic terms. *Chemical Reviews*, *101*(3), 619–672. doi:10.1021/cr0000067 PMID:11712499

Hansch, C., & Leo, A. (1979). *Substituent constants for correlation analysis in chemistry and biology*. New York: John Wiley & Sons.

Hansch, C., & Leo, A. (1995). *Fundamentals and applications in chemistry and biochemisry* (Q. S. A. R. Exploring, Ed.). Washington, DC: American Chemical Society.

Hansch, C., & Selassie, C. (2007). *Quantitative structure-activity relationship-a historical perspective and the future*. Oxford, UK: Elsevier.

Hartmann, C., Antes, I., & Lengauer, T. (2009). Docking and scoring with alternative side-chain conformations. *Proteins*, *74*(3), 712–726. doi:10.1002/prot.22189 PMID:18704939

Hill, M. D., Jouppi, N. P., Sohi, & Gurindar. (1999). *Readings in computer architecture*. Academic Press.

Hoffman, A. R., et al. (1990). *Supercomputers: Directions in technology and applications*. National Academies. Retrieved June 17, 2014 from http://www.intechopen.com/books/protein-engineering-technology-and-application/protein-protein-and-protein-ligand-docking

Irwin, J. J., & Shoichet, B. K. (2005). ZINC — A free database of commercially available compounds for virtual screening. *Journal of Chemical Information and Modeling*, *45*(1), 177–182. doi:10.1021/ci049714+ PMID:15667143

Jayaram, B. (2011). *DNA ligand docking*. Retrieved June 16, 2014 from http://www.scfbio-iitd.res.in/dock/dnadock.jsp

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Jayaram, B. (2012). *BAPPL server*. Retrieved June 16, 2014 from http://www.scfbio-iitd.res.in/software/drugdesign/bappl.jsp

Jayaram, B. (2013). *ParDOCK - Automated server for protein ligand docking*. Retrieved June 16, 2014 from http://www.scfbio-iitd.res.in/dock/pardock.jsp

Joachimiak, A. (2009). High-throughput crystallography for structural genomics. *Current Opinion in Structural Biology*, *19*(5), 573–584. doi:10.1016/j.sbi.2009.08.002 PMID:19765976

Jorgensen, W. L. (1991). Rusting of the lock and key model for protein-ligand binding. *Science*, *254*(5034), 954–955. doi:10.1126/science.1719636 PMID:1719636

Kearsley, S. K., Underwood, D. J., Sheridan, R. P., & Miller, M. D. (1994). Flexibases: A way to enhance the use of molecular docking methods. *Journal of Computer-Aided Molecular Design*, *8*(5), 565–582. doi:10.1007/BF00123666 PMID:7876901

Kitchen, D. B., Decornez, H., Furr, J. R., & Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nature Reviews. Drug Discovery*, *3*(11), 935–949. doi:10.1038/nrd1549 PMID:15520816

Klebe, G., & Mietzner, T. (1994). A fast and efficient method to generate biologically relevant conformations. *Journal of Computer-Aided Molecular Design*, *8*(5), 583–606. doi:10.1007/BF00123667 PMID:7876902

Krivobok, S., Kuony, S., Meyer, C., Louwagie, M., & Wilson, J. C. (2003). Identification of pyrene-induced proteins in *Mycobacterium* spp. strain 6PY1: Evidence for two ring-hydroxylating dioxygenas-es. *Journal of Bacteriology*, *185*(13), 3828–3841. doi:10.1128/JB.185.13.3828-3841.2003 PMID:12813077

Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R., & Ferrin, T. E. (1982). A geometric approach to macromolecule–ligand interactions. *Journal of Molecular Biology*, *161*(2), 269–1288. doi:10.1016/0022-2836(82)90153-X PMID:7154081

Labana, S., Pandey, G., Paul, D., Sharma, N. K., Basu, A., & Jain, R. K. (2005). Plot and field studies on bioremediation of *p*-nitrophenol contaminated soil using *Arthrobacter protophormiae* RKJ100. *Environmental Science & Technology*, *39*(9), 3330–3337. doi:10.1021/es0489801 PMID:15926586

Labana, S., Singh, O. V., Basu, A., Pandey, G., & Jain, R. K. (2005). A microcosm study on bioremediation of p-nitrophenol-contaminated soil using *Arthrobacter protophormiae* RKJ100. *Applied Microbiology and Biotechnology*, *68*(3), 417–424. doi:10.1007/s00253-005-1926-1 PMID:15806356

Lengauer, T., & Rarey, M. (1996). Computational methods for bimolecular docking. *Current Opinion in Structural Biology*, *6*(3), 402–406. doi:10.1016/S0959-440X(96)80061-3 PMID:8804827

Lipman, D. J., & Pearson, W. R. (1985). Rapid and sensitive protein similarity searches. *Science*, *227*(4693), 1435–1441. doi:10.1126/science.2983426 PMID:2983426

List of Sequence Alignment Software. (2014). Retrieved June 17, 2014 from http://en.wikipedia.org/wiki/List_of_sequence_alignment_software

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Livingstone, D. J. (2000). The characterization of chemical structures using molecular properties. A survey. *Journal of Chemical Information and Computer Sciences*, *40*(2), 195–209. doi:10.1021/ci990162i PMID:10761119

Lovley, D. R. (2003). Cleaning up with genomic: Applying molecular biology to bioremediation. *Nature Reviews. Microbiology*, *1*(1), 35–44. doi:10.1038/nrmicro731 PMID:15040178

Marmur, J., & Doty, P. (1961). Thermal renaturation of deoxyribonucleic acids. *Journal of Molecular Biology*, *3*(5), 585–594. doi:10.1016/S0022-2836(61)80023-5 PMID:14470100

Matsuzaki, Y., Uchikoga, N., Ohue, M., Shimoda, T., Sato, T., Ishida, T., & Akiyama, Y. (2013). MEGADOCK 3.0: A high-performance protein-protein interaction prediction software using hybrid parallel computing for petascale supercomputing environments. [MEGADOCK 3.0: a high-performance protein-protein interaction prediction software using hybrid parallel computing for petascale supercomputing environments.] *Pubmed, 8*(18). doi:10.1186/1751-0473-8-18

Mendoza-Barrera, C., Hernandez-Santoyo, A., Tenorio-Barajas, A. Y., & Altuzar, V. (2013). *Protein-protein and protein-ligand docking*. Retrieved June 17, 2014 from http://www.intechopen.com/books/protein-engineering-technology-and-application/protein-protein-and-protein-ligand-docking

Meng, E. C., Shoichet, B. K., & Kuntz, I. D. (2004). Automated docking with grid-based energy evaluation. *Journal of Computational Chemistry*, *13*(4), 505–524. doi:10.1002/jcc.540130412

Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., & Olson, A. J. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry*, *19*(14), 1639–1662. doi:10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B

O'Neal, T. (2014). *Novel analyses improve identification of cancer-associated genes*. Retrieved June 16, 2014 from http://www.supercomputingonline.com/latest/topics/this-month/57971-novel-analyses-improve-identification-of-cancer-associated-genes

Pandey, G., Paul, D., & Jain, K. (2003). Branching of *o*-nitrobenzoate degradation pathway in *Arthrobacter protophormiae* RKJ100: Identification of new intermediates. *FEMS Microbiology Letters*, *229*(2), 231–236. doi:10.1016/S0378-1097(03)00844-9 PMID:14680704

Pandey, U. B., & Nichols, C. D. (2011). Human disease models in Drosophila melanogaster and the role of the fly in therapeutic drug discovery. *Pharmacological Reviews*, *63*(2), 2411–2436. doi:10.1124/pr.110.003293 PMID:21415126

Pearson, W. R., & Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences of the United States of America*, *85*(8), 2444–2448. doi:10.1073/pnas.85.8.2444 PMID:3162770

Ponder, B. A. (2001). Cancer genetics. *Nature*, *411*(6835), 336–341. doi:10.1038/35077207 PMID:11357140

Prodan, R. (2007). Grid computing: Experiment management, tool integration, and scientific workflows. Fahringer & Thomas.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Richmond, C. S., Glasner, J. D., Mau, R., Jin, H., & Blattner, F. R. (1999). Genome-wide expression profiling in *Escherichia coli* K-12. *Nucleic Acids Research*, *27*(19), 3821–3835. doi:10.1093/nar/27.19.3821 PMID:10481021

Ripphausen, P., Nisius, B., Peltason, L., Bajorath, J., Quo, & Vadis. (2010). Virtual screening? A comprehensive survey of prospective applications. *Journal of Medicinal Chemistry*, *53*, 8461-8467.

Rise, M. L., Jones, S. R., Brown, G. D., Von Schalburg, K. R., Davidson, W. S., & Koop, B. F. (2004). Microarray analyses identify molecular biomarkers of Atlantic salmon macrophage and hematopoietic kidney response to *Piscirickettsia salmonis* infection. *Physiological Genomics*, *20*(1), 21–35. doi:10.1152/physiolgenomics.00036.2004 PMID:15454580

Rowley, J. D. (1973). A new consistent chromosomal abnormality in chronic myelogenous leukemia identified by quina-crine fluorescence and Giemsa staining. *Nature*, *243*(5405), 290–293. doi:10.1038/243290a0 PMID:4126434

Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P. O., & Davis, R. W. (1996). Parallel human genome analysis: Microarray-based expression of 1000 genes. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(20), 10539–11286. doi:10.1073/pnas.93.20.10614 PMID:8855227

Schut, G. J., Zhou, J., & Adams, M. W. (2001). DNA microarray analysis of the hyperthermophilic archaeon *Pyrococcus furi-osus*: Evidence for a new type of sulfur-reducing enzyme complex. *Journal of Bacteriology*, *183*(24), 7027–7036. doi:10.1128/JB.183.24.7027-7036.2001 PMID:11717259

Shalon, D., Smith, S. J., & Brown, P. O. (1996). A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Research*, *6*(7), 639–645. doi:10.1101/gr.6.7.639 PMID:8796352

Singh, A. et al. (2013). A review on DNA microarray technology. *International Journal of Current Research and Review*, *05*(22), 5.

Structure Based drug Design and Molecular Modeling. (n.d.). Retrieved June 16, 2014 from http://www.imb-jena.de/~rake/Bioinformatics_WEB/dd_tools.html

Sundberg, S. A., Chow, A., Nikiforov, T., & Wada, G. (2001). Microchip-based systems for biomedical and pharmaceutical analysis. *European Journal of Pharmaceutical Sciences*, *14*(1), 1–12. doi:10.1016/S0928-0987(01)00153-1 PMID:11457644

Superimpose. (2010). Retrieved June 16, 2014 from http://www.scfbio-iitd.res.in/software/utility/Superimpose.jsp

Tao, H., Bausch, C., Richmond, C., Blattner, F. R., & Conway, T. (1999). Functional genomics: Expression analysis of *Escherichia coli* growing on minimal and rich media. *Journal of Bacteriology*, *181*, 6425–6440. PMID:10515934

Taylor, P., Blackburn, E., Sheng, Y. G., Harding, S., Hsin, K. Y., & Kan, D. et al. (2008). Ligand discovery and virtual screening using the program LIDAEUS. *British Journal of Pharmacology*, *153*, 555–567. PMID:18037921

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Taylor, R. D., Jewsbury, P. J., & Essex, J. W. (2003). FDS: Flexible ligand and receptor docking with a continuum solvent model and soft-core energy function. *Journal of Computational Chemistry*, *24*(13), 1637–1656. doi:10.1002/jcc.10295 PMID:12926007

Totrov, M., & Abagyan, R. (2008). Flexible ligand docking to multiple receptor conformations: A practical alternative. *Current Opinion in Structural Biology*, *18*(2), 178–184. doi:10.1016/j.sbi.2008.01.004 PMID:18302984

Valafar, F. (2002). Pattern recognition techniques in microarray data analysis: A survey. *Annals of the New York Academy of Sciences*, *980*(1), 41–64. doi:10.1111/j.1749-6632.2002.tb04888.x PMID:12594081

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., & Sutton, G. G. et al. (2001). The sequence of the human genome. *Science*, *291*(5507), 1304–1351. doi:10.1126/science.1058040 PMID:11181995

Wall, M., Rechtsteiner, A., & Rocha, L. (n.d.). *Singular value decomposition and principal component analysis*. Retrieved June 16, 2014 from http://public.lanl.gov/mewall/kluwer2002.html

Wang, Q., & Pang, Y. P. (2007). Preference of small molecules for local minimum conformations when binding to proteins. *PLoS ONE*, *2*(9), e820. doi:10.1371/journal.pone.0000820

Wei, B. Q., Weaver, L. H., Ferrari, A. M., Matthews, B. W., & Shoichet, B. K. (2004). Testing a flexible-receptor docking algorithm in a model binding site. *Journal of Molecular Biology*, *337*(5), 1161–1182. doi:10.1016/j.jmb.2004.02.015 PMID:15046985

Wei, H., Wang, C. H., Du, Q. S., Meng, J., & Chou, K.-C. (2009). Investigation into adamantane-based M2 inhibitors with FB-QSAR. *Journal of Medicinal Chemistry*, *5*(4), 305–317. doi:10.2174/157340609788681430 PMID:19689387

Willighagen, E. (2011). *An ontology for QSAR and cheminformatics*. Retrieved June 16, 2014 from http://chem-bla-ics.blogspot.in/2011/10/ontology-for-qsar-and-cheminformatics.html

Wilson, M., DeRisi, J., Kristensen, H. H., Imboden, P., Rane, S., Brown, P. O., & Schoolnik, G. K. (1999). Exploring drug-induced alterations in gene expression in *Mycobacterium tuberculosis* by microarray hybridization. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(22), 12833–12838. doi:10.1073/pnas.96.22.12833 PMID:10536008

Yang, Y. (2006). *Bioinformatics resources for microarray data analysis, protein function/structure prediction, and protein-protein interaction*. Retrieved June 16, 2014 from http://dragon.bio.purdue.edu/bioinfolinks/

Ye, R. W., Tao, W., Bedzyk, L., Young, T., Chen, M., & Li, L. (2000). Global gene expression profiles of *Bacillus subtilis* grown under anaerobic conditions. *Journal of Bacteriology*, *182*(16), 4458–4465. doi:10.1128/JB.182.16.4458-4465.2000 PMID:10913079

Yoshida, K., Kobayashi, K., Miwa, Y., Kang, C. M., Matsunaga, M., & Yamaguchi, H. et al. (2001). Combined transcriptome and proteome analysis as a powerful approach to study genes under glucose repression in *Bacillus subtilis*. *Nucleic Acids Research*, *29*(3), 683–692. doi:10.1093/nar/29.3.683 PMID:11160890

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Young, R. A. (2000). Biomedical discovery with DNA arrays. *Cell*, *102*(1), 9–15. doi:10.1016/S0092-8674(00)00005-2 PMID:10929708

Zhang, C., Li, P., Rajendran, A., Deng, Y., & Chen, D. (2006). *Parallelization of multicategory support vector machines (PMC-SVM) for classifying microarray data*. Retrieved June 16, 2014 from http://link.springer.com/article/10.1186%2F1471-2105-7-S4-S15

Zsoldos, R. Z., Simon, D., Sadjad, A. S. B., & Johnson, A. P. (2007). eHiTS: A new fast, exhaustive flexible ligand docking system. *Journal of Molecular Graphics & Modelling*, *26*(1), 198–212.

KEY TERMS AND DEFINITIONS

Data Analysis: Analysis and isolation of information from the provided database using supercomputers. These data may be protein, DNA, Genome, Proteome etc.

Docking: Close interaction studies between protein and ligand or between protein and protein. It involves formation of bonds between amino acids and between amino acids and drug molecule.

Genomics: Study of genes and genome by utilizing computational as well as experimental knowledge.

Homology Modeling: 3D structure prediction of protein on the basis of similarity and identity between two or more sequences from the databases.

Ligand: Small Drug or interacting molecule which are able to dock with receptor (Protein). These ligand are able to perform specific functions after interaction with receptors.

Protein-Protein Docking: Interaction studies between two proteins like one receptor and another protein or between peptides.

Quantitative Structure Activity Relationship (QSAR): Used to define the function of the molecules or atoms on the basis of their structure.

Sequence Analysis: Analysis of protein or DNA nucleotide sequence and identification of sequence conservation, hot spots, active site, identity, homology and phylogeny among the sequences.

Supercomputers: Introduced in 1960s and it is a computer with high processing capacity particularly speed of calculation.

Virtual Screening: Screening or filtering of molecules on the basis of similarity of their structure. It is a database search method.

This work was previously published in Research and Applications in Global Supercomputing edited by Richard S. Segall, Jeffrey S. Cook, and Qingyu Zhang, pages 223-241, copyright year 2015 by Information Science Reference (an imprint of IGI Global).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 22
QSAR Models towards Cholinesterase Inhibitors for the Treatment of Alzheimer's Disease

C. Gopi Mohan
Amrita Institute of Medical Sciences and Research Centre, India

Shikhar Gupta
National Institute of Pharmaceutical Education and Research, India

ABSTRACT

Alzheimer's Disease (AD) is a multifactorial neurological syndrome with the combination of aging, genetic, and environmental factors triggering the pathological decline. Interestingly, the importance of the Acetylcholinesterase (AChE) enzyme has increased due to its involvement in the β-amyloid peptide fibril formation during AD pathogenesis. In silico technique, QSAR has proven its usefulness in pharmaceutical research for the design/optimization of new chemical entities. Further, QSAR method advanced the scope of rational drug design and the search for the mechanism of drug action. It is a well-established fact that the chemical and pharmaceutical effects of a compound are closely related to its physico-chemical properties, which can be calculated by various methods from the compound structure. This chapter focuses on different Quantitative Structure-Activity Relationship (QSAR) studies carried out for a variety of cholinesterase inhibitors for the treatment of AD. These predictive models will be potentially used for further designing better and safer drugs against AD.

INTRODUCTION

Alzheimer's Disease (AD) is a multi-factorial syndrome with the combination of aging, genetic, non-genetic causes, and environmental factors triggering the pathological decline (Butters, Deliss & Lucas, 1995). It is a most common form of dementia, with chronic, irreversible and progressive neurodegenerative disorder.

DOI: 10.4018/978-1-5225-0549-5.ch022

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

AD usually begins after the age of 60, and the risk increases as the age progresses. Younger people in their 30s to 50s may get AD, but it is rare. Approximately, 10% of all the cases of AD are believed to be hereditary in nature. In familial cases, symptoms usually appear within the age range of 30-60 years and known as early-onset AD. This type of AD results from specific genetic mutation of the individual. The late-onset sporadic AD, representing 90% of patients result from multifactorial environmental factors and genetic events, caused due to the inheritance of the apolipoprotein Eε4 allele and other acting polymorphic genes (Rosenberg, 2000). The non-genetic factors are also playing an important role in AD, as the only one third of the identical twins is concordant of the disease (Jin, Gatz, Johansson & Pedersen, 2004). Still the role of the environmental risk factors has remained mysterious, but progress in finding genes causes of AD, as well as increasing risk for it has been firm and impressive.

Several evidences showed that reactive oxygen species (ROS) are involved in most of the neurodegenerative pathologies. ROS is the most dangerous reactive species, and which can spoil most of the biological molecules. Hence, it is crucial to maintain the oxidative balance and control in the brain. This is known as the oxygen paradox-oxygen, which is an absolute necessity for our energy-economical aerobic life style. However, it is a potential toxin and our brain is most aerobically active organ, which required ~20% of total oxygen in a resting individual. Normally, these free radicals quickly detoxified by the body's defense mechanism and firmly regulated by antioxidants. Hence, modifications in normal oxidative metabolism as observed in AD brain suggest that oxidative stress has a crucial role in AD pathogenesis. In general, chemical origin of the bulk of ROS is the reaction of the molecular oxygen with the redox active metals Fe and Cu.

Two distinct histological changes occur in the nerve cells of Alzheimer brain are the formation of extracellular amyloid ('senile') plaques developed between neurons and intracellular neurofibrillary tangles developed within neurons, which lead to neurotoxicity. Plaques, which are composed of β-amyloid polypeptides (Aβ) formed by the mutations in the amyloid precursor protein (APP) gene on chromosome 21q and of the presenilin 1 (PS1) and presenilin 2 (PS2) genes on chromosomes 14q and 1q, respectively (Tanzi, Kovacs, Kim, Moir, Guenette, & Wasco, 1996). Amyloidogenesis are responsible for almost one half of the early-onset forms of autosomal dominant inherited disease. An increased synthesis of Aβ in the AD brain is a central point in the amyloid hypothesis and suggests that increased amyloidogenesis and/or decreased amyloid clearance with increased amyloid fibrillation are primarily causal of the pathogenesis of AD (Naslund, Haroutunian, Mohs, Davis, Davies, Greengard, & Buxbaum, 2000). Neurofibrillary tangles are secondary important histological abnormalities, which are composed of hyperphosphorylated tau protein and links together to form filaments. Increased density of tau deposition within neurons in the brain facilitates Aβ toxicity. APP processing involves 3 classes of enzymes: α-, β-, and γ-secretase (Mullan, Crawford, Axelman, Houlden, Lilius, Winblad, & Lannfelt, 1992). APP is first enzymatically cleaved by γ- or β-secretase, and which in turn was cleaved by γ-secretase. The segment of the molecule produced by α- γ cleavage yielded a soluble fragment and a self-aggregating fragment (β amyloid 40–42) from the portion of the molecule produced by the β-γ cleavage.

Cholinergic hypothesis targeting acetylcholinesterase (AChE) enzyme are one of the major therapeutic strategies adopted for symptomatic relief on AD (Bartus, Dean, Beer & Lippa, 1982). It is a substrate-specific essential enzyme in the family of serine hydrolases, and which degrades the neurotransmitter acetylcholine (ACh) in the nerve synapses. An optimum level of ACh should be maintained in the hippocampus and the cortex region of the brain for its proper function (Stahl, 1999). This hypothesis is proven to be successful today by the effective use of cholinesterase inhibitors, to augment the surviving cholinergic activity for the treatment of AD.

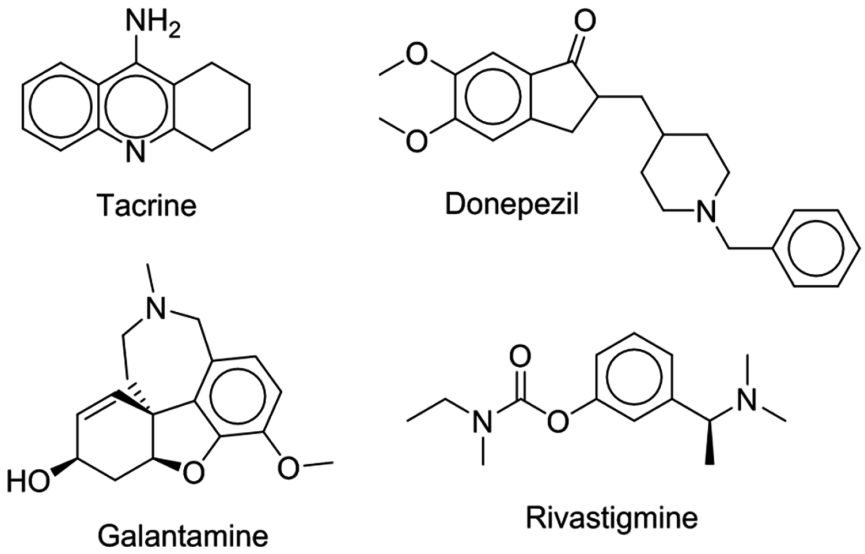
EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Cholinesterase inhibitors such as tacrine, rivastigmine, donepezil and galantamine, which block the breakdown of ACh has been used for the treatment of mild to moderate AD (Mayeux & Sano, 1999; Silmana & Sussman, 2008, Mirjana, 2013). All of these drugs shown in Figure 1 are reversible AChE inhibitors (AChEIs), and can restore the level of ACh in the brain of AD patients, and which plays a key role in memory and cognition (Sugimoto, 2000; Colombers, 2004).

Hence, it is not surprising that AChEIs have shown better benefits in the treatment of AD than other therapeutic strategies. Recently, interference on amyloid beta (Aβ) aggregation by AChEIs directed to the development of a novel class of AChEIs. It was also experimentally observed that AChEIs moderates the rate of amyloid deposition (Inestrosa, Dinamarca, Alvarez, & Center, 2008). The bifunctional role, anti-aggregating Aβ (characteristic pathological role in AD) and anti-cholinesterase activity is mainly contributed by these AChEIs with interactions on this enzyme (De Ferrari, Canales, Shin, Weiner, Silman, & Inestrosa, 2001). Therefore, identifying selective AChEIs are highly demanding for CNS penetration and good pharmacokinetic properties for the treatment of AD.

Many studies have focused on synthesizing and evaluating new AChEIs for the improvement of symptoms of AD. One of the major strategies has been the use of *in silico* techniques such as QSAR modeling, molecular docking, pharmacophore mapping, virtual screening etc. which has proven their usefulness in pharmaceutical research for the selection/identification and/or design/optimization of new chemical entities. QSAR is one of the most important areas in Chemoinformatics, and its advances have widened the scope of rational drug design and the search for the mechanism of drug action (Hansch, Hoekman, Leo, Weininger, & Selassie, 2002; Garg, Gandhi & Mohan, 2008). It is a well-established fact that the chemical and pharmacological effects of a compound are closely related to its physico-chemical properties. This in turn been calculated by various methods from the molecular structure (Gupta, 2007). In addition, the *in silico* methods can expand screening for compounds that do not exist physically in the chemical collections, therefore compensating for some of the most important limitations of the heavy experimental measurements which include high-throughput *in vitro* screening methods.

Figure 1. Acetylcholinesterase inhibitors approved by FDA for the treatment of AD



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

ACETYLCHOLINESTERASE ENZYME

Dale in 1914 postulated that the physiological role of ACh at cholinergic synapses is terminated by an enzyme system. It was not until 1926 that Loewi and Navratil experimentally established AChE existence and in 1932 Stedman arranged a crude extract of an "ACh-splitting" enzyme from horse serum, which he called "cholinesterase". The enzyme AChE (ACh hydrolyser, EC 3.1.1.7) is a molecule 32 KDa molecular weight. The principal biological role of AChE is the termination of impulse transmission at cholinergic synapses by rapid hydrolysis of the neurotransmitter ACh to produce choline and an acetate group (Barnard 1974), and is presented in Figure 2.

AChE is mainly found at neuromuscular junctions and cholinergic synapses in the central nervous system, where its activity serves to terminate the synaptic transmission. AChE has a very high catalytic activity, and each molecule of AChE degrades about 5000 molecules of ACh per second. The choline produced by the action of AChE is recycled - it is transported through reuptake into nerve terminals, where it is used to synthesize new ACh molecules (Quinn, 1987). AChE exists in multiple molecular forms, which possess similar catalytic properties, but differ in their oligomeric assembly and mode of attachment to the cell surface. Knowledge of the three dimensional structure of AChE is essential for understanding its remarkable catalytic efficacy.

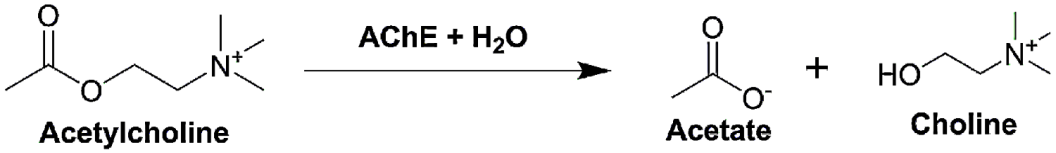
General Structure of AChE

Ever since the crystal structure of *Torpedo californica* acetylcholinesterase (*Tc*AChE) was solved in 1991, it opened up new horizons in the understanding of catalytic mechanisms and mode of action of inhibitors with this enzyme (Sussman *et al.* 1991). *Tc*AChE belongs to the class of α/β proteins (Levitt, & Chothia, 1976) and consists of 12-stranded central mixed β-sheet surrounded by 14 α-helices as shown in Figure 3. The enzyme has a β-sheet platform that bears the catalytic machinery and is rather similar in all members of the family. Indeed, the three members of the catalytic triad appeared in the same order along the polypeptide chain in all α/β hydrolase enzymes (Ollis *et al.* 1992). The α-helices and loops are then ascribed by the task of handling the specificity element, substrates of different members of the family being very varied. When *Tc*AChE structure was determined, it initially appeared to be an entirely new fold. However, it folds soon turned out to be strikingly similar to other proteins whose structures had been solved almost the same time.

The Active-Site Gorge and the Catalytic Triad

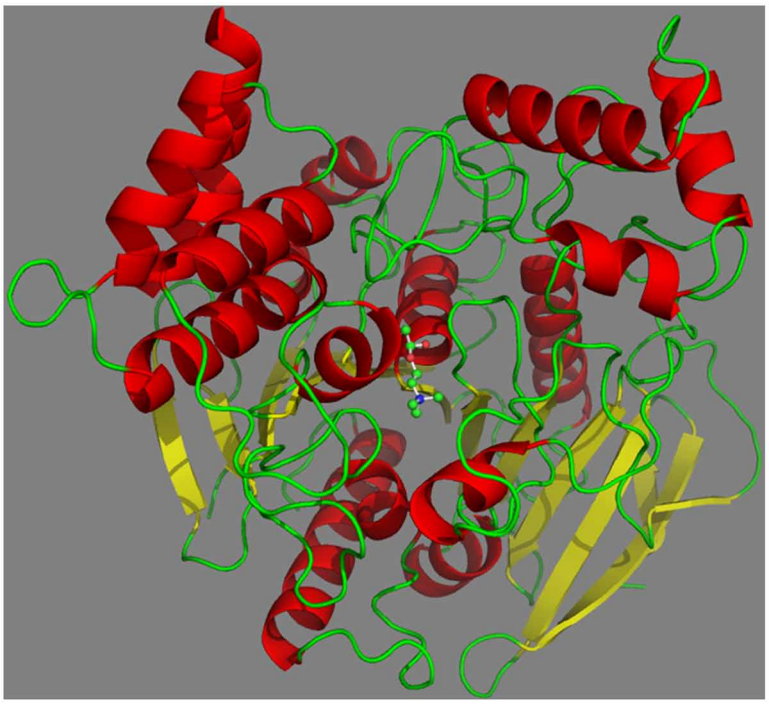
The most remarkable feature of the structure of *Tc*AChE, is a deep and narrow gorge, about 20 Å long. This region, which penetrates more than half-way into the enzyme, and lined with the rings of 14 con-

Figure 2. Enzymatic hydrolysis of ACh by AChE



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 3. Schematic ribbon diagram of the 3D structure of TcAChE monomer; 11-standard β-sheets (light gray) surrounded by 15 α-helices (dark black) and 3 β-sheets without any hydrogen bond to the central sheet; ACh in the binding site has been rendered in ball and stick format.

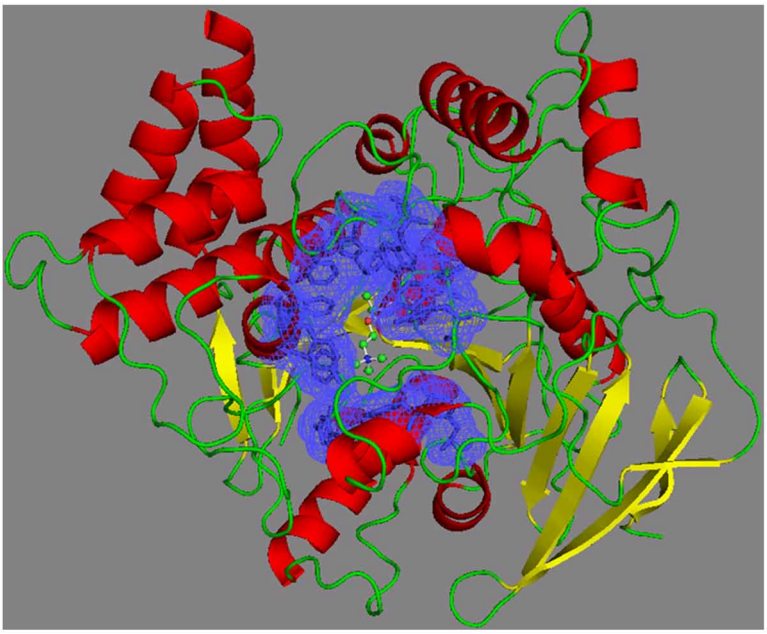


served aromatic residues is presented in Figure 4. The existence of a catalytic triad in AChE had previously been the subject of controversy. As revealed by the crystallographic structure of AChE and its inhibitor complexes, the AChE active site contains a catalytic triad (S200, H440, E327) located at the bottom of a deep and narrow gorge as shown in Figure 5. Early kinetic studies indicated that the catalytic site (CS) of AChE contains two subsites, the 'esteratic' and 'anionic' subsites (AS) (Rosenberry, 1975). These two subsites correspond to the catalytic machinery and the choline-binding pocket. The 'esteratic' subsite was believed to resemble the catalytic subsites of other serine hydrolases (Froede, & Wilson, 1971). The AS, located near the bottom of the cavity (Kryger, Silman, & Sussman, 1999) and interacts with the charged quaternary moiety of choline group of ACh was believed to be the binding site both for quaternary ligands, such as edrophonium and N-methylacridinium. They act as competitive inhibitors, and for quaternary oximes, which serve as effective reactivators of organophosphate-inhibited AChE (Mooser, & Sigman, 1974).

Both chemical modification and spectroscopic studies supported the presence of aromatic residues in the active site of AChE (Fuchs, Gurari, & Silman, 1974). These studies implicated a histidine (H440) residue in the active site (Goeldner, & Hirth, 1980). S200-Oγ, which can be seen looking down the gorge from the surface of the enzyme, is about 4 Å above the base of the gorge. The rings of 14 aromatic residues contributed a substantial portion (~40%) of the surface of the gorge as shown in Figure 4. These residues, and their flanking sequences, which are highly conserved in AChEIs from different species was N66 and I444. They are synthesized on the first exon, which codes for residues 1–480. It should be noted

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 4. 3D structure of TcAChE displayed as a ribbon diagram; the 14 conserved aromatic residues are shown as sticks and mesh. A model of the substrate, ACh, bound in the active site, is shown at the bottom of the active-site gorge, in ball and stick format.



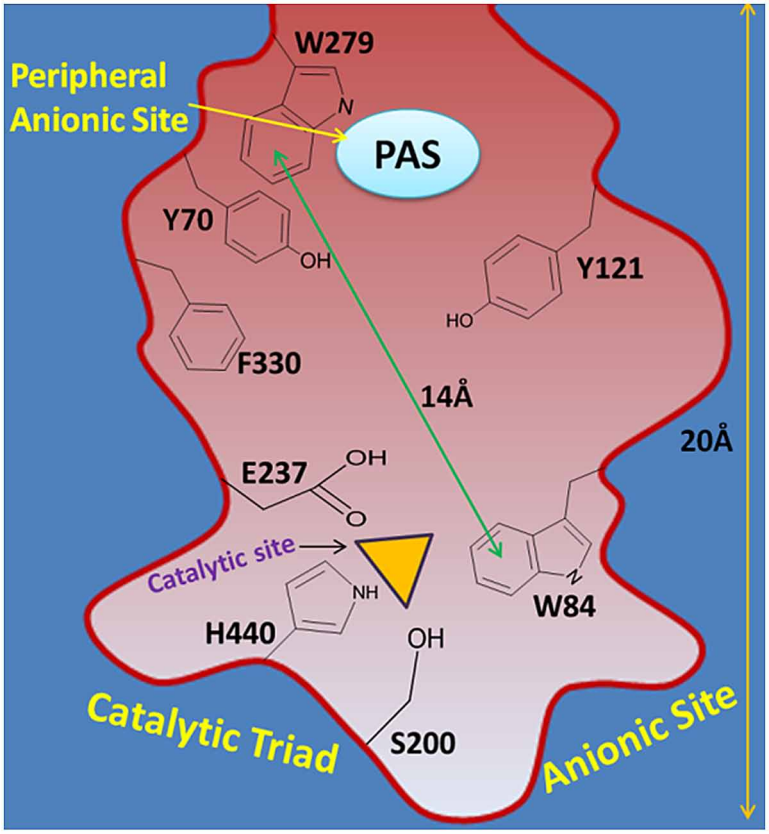
that the gorge contains only a few acidic residues, which include D285 and E273 at the very top, D72, hydrogen-bonded to Y334, about half-way down, and E199, near the base. The presence of a tryptophan in the active site of AChE was predicted by spectroscopic and chemical modification studies (Maulet, Camp, Gibney, Rachinsky, Ekströ, & Taylor, 1990). The subsequent affinity labeling study of Weise *et al.* indeed identified W84 as part of the putative 'anionic' (choline) binding site. The hydroxyl groups of Y121 (half-way up) and Y130 (at the bottom) both points into the gorge (Weise, Kreienkamp, Raba, Pedak, Aaviksaar, & Hucho, 1990). The high aromatic content of the walls and base of the active-site gorge, together with its dimensions, may help explain why biochemical studies revealed a variety of hydrophobic and 'anionic' binding sites distinct from, or overlapping with the active site.

Peripheral Anionic Site

Besides the catalytic site, the peripheral anionic site (PAS) of AChE is placed on the surface of the protein. W279 and Y70 were introduced as the residues of PAS and is presented in Figure 5. Furthermore, two sets of peptide sequences, residues 251-264 and 270-278, respectively, contribute to the PAS in *Tc*AChE for ACh and other quaternary ligands. These two neighboring peptide sequences on the surface of the protein are both close to the rim of the gorge. Thus, the ligand association with the peripheral site may prevent access of the substrate to the gorge by physical hindrance to restrict entry to the gorge by an allosteric mechanism, in which the active center conformation is altered (Taylor, & Radic, 1994). This blockade can have functional significance. In particular, during ACh hydrolysis, an incoming ACh molecule bound at PAS blocks the exit of a choline molecule generated in the active site, giving rise

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 5. Schematic view of the active-site gorge of TcAChE
(Gupta S, 2012).



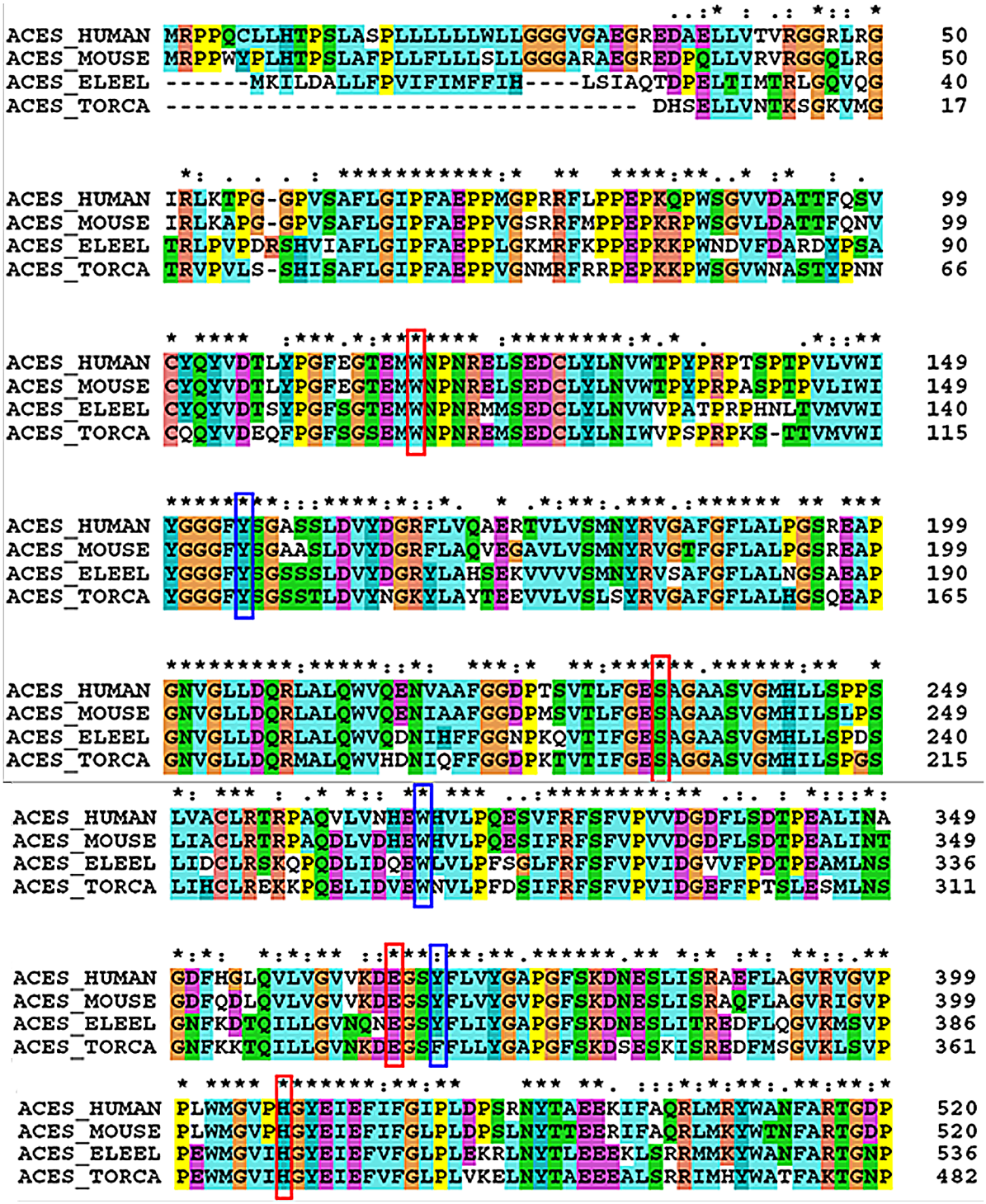
to the substrate inhibition. Some ligands may be too bulky to penetrate it, but can partially block its entrance. Certain elongated bis-quaternary compounds, many of which serve as potent inhibitors, may attach at one extremity to the PAS, and at other end to any one of the various aromatic residues lining the walls of the gorge, in certain cases spanning the gorge completely (Harel *et al.* 1993). Recently, evidence was presented that AChE accelerates assembly of Aβ peptides into amyloid fibrils with the involvement of PAS.

Comparison of Human hAChE and *Tc*AChE

Among the available crystal structures of the AChE enzyme, docking studies were performed on the donepezil bound crystal structure of *Tc*AChE (PDB ID: 1EVE) (Kryger, Silman, & Sussman, 1999). We have selected this crystal structure by considering that the size and shape complementarity as well as the dual binding site nature of the donepezil compound was similar to our training set lead compound **CD1**. *Tc*AChE has almost identical amino acid residues with the *human* AChE (hAChE) (PDB code: 1B41) at both the CS and PAS, apart from the substitution of F330 (*Tc*) with Y337 (*human*) (Kryger, Harel, & Shafferman, 2000). The multiple sequence alignment (MSA) of AChE sequences from *Homo sapiens*, *Mus musculus*, *Electric eel*, and *Torpedo californica* is shown in Figure 6. MSA shows that the active

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 6. The multiple sequence alignment of AChE sequences from Homo sapiens, Mus musculus, Electric eel, and Torpedo californica, boxes showing CS and PAS residues



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

site of AChE is very well conserved across species. hAChE and mAChE sequences exhibit 88% identity and 97% homology, while with hAChE and *Tc*AChE sequences exhibit 64% identity and 88% homology.

Mechanism of Action

AChE is the enzyme that catalyzes the breakdown of the cationic neurotransmitter, ACh: a neurotransmitter important in cognitive function, in the nerve synapses. The catalytic triad has been termed as "charge relay system". In *Tc*AChE, the triad includes S200, the imidazole ring of H440 and the carboxylic group of E327. It acts through an acylation and deacylation process (Wang, Jiang, Chen, Ji, & Ye, 1999). In the active site S200 and H440 are involved in reaction with ACh. During binding of ACh to AChE, the acyl-enzyme is produced after proton transformation from S200 to the imidazole group of H440, and then oxygen of S200 attacks to the substrate (ACh), this part is acylation and then the acylated enzyme complex is finally rapidly hydrolyzed with the waters of binding pocket, regenerating active enzyme by releasing acetic acid. During the acylation process, the cation-π interaction takes place between the positively charged nitrogen in ACh and W84.

Furthermore, the component of oxyanion holes (G118, G119 and A201) forms the hydrogen bond to the carbonyl group of ACh. In the acylation process, both the cation interaction and H-bonds exist during the reaction. However, these two kinds of interactions indirectly affect the proton transfer from S200 to H440 and the nucleophilic attack of the oxygen atom of S200 to the carbonyl of ACh.

BACKGROUND

AChEIs can be classified into different groups, based on their mode of interaction with the AChE enzyme. Inhibition can be either reversible, by competitively preventing the substrate from reaching the active site; pseudo-irreversible, by covalent reaction with the active site serine, inactivating the catalytic ability of the enzyme or irreversible. Competitive inhibitors act by blocking substrate at the active site, non-competitive inhibitors by binding to the PAS.

Synaptic Cholinergic Drugs

In tissues, the most abundant of cholinesterase types are AChE and BuChE. As it was described earlier, AChE is the predominant enzyme in the brain. Therefore, inhibition of AChE causes more bioavailability of ACh at the synaptic area and consequently improving neurotransmission process. Principally, this method is most useful for the treatment of patients with undamaged presynaptic neurons. They are still active in synthesizing and releasing ACh. It works in the early stages of AD and loses effectiveness after usage in a period of time. The activation of M2 muscarinic receptors that leads to inhibition of presynaptic release of ACh might decrease the efficacy of AChEIs, through the counteracting effect. Despite this AChEIs have shown suitable therapeutic effects for patients having AD. The only drugs currently accepted to treat the AD are AChEIs (i.e. tacrine, donepezil, rivastigmine and galanthamine). Based on the mechanism of action in AChE, different sort of AChEIs have been designed and classified as pseudo-irreversible, irreversible, transition state analogue and reversible inhibitors respectively (Camps, & Munoz-Torrero, 2002).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Irreversible AChEIs (Organophosphorus Compounds)

Organophosphates compounds such as di-isopropyl fluorophosphates (DFP) are very potent inhibitors of AChE and are used as agricultural insecticides or as nerve gases in chemical warfare. One of the representatives of this group is metrifonate (Figure 7), which irreversibly phosphorylates the enzyme. Although its efficacy was acceptable, its application as a drug was withdrawn due to muscle weakness and respiratory problem. These compounds react with the active site serine, forming a very stable covalent phosphoryl-enzyme complex.

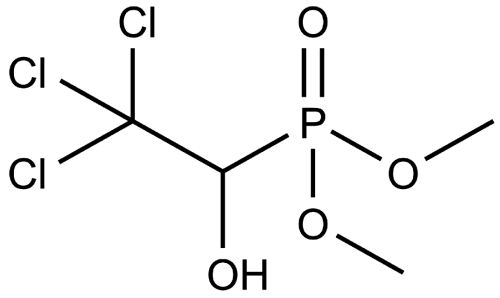
Pseudo-Irreversible AChEIs

Pseudo-irreversible AChEIs includes the compounds having carbamates functional group. They are carbamylated by a catalytic triad of the AChE binding site. The rate of hydrolization of their carbamoylated complex with S200 is slower than the rate of hydrolization of ACh-AChE complex. Physostigmine was the first AChEI of this class that was studied for the treatment of AD, but it was rejected due to the lack of efficacy resulting from its short half-life and variable bioavailability. To improve its potency, several analogues of that with more lipophilic side chains have been designed. Rivastigmine (Exelon®) is a miotine derivative. It is a carbamic AChEI with a short half-life and duration of action of about 10 hours. This drug inhibits AChE by carbamoylating the serine residue of the catalytic triad in a pseudo-irreversible manner. However, it shows a good combination in brain selectivity, long duration *in vivo* activity and its good neuroprotective property, which caused it to be accepted as a drug for AD treatment. Eptastigmine, a more recent physostigmine analogue, is the heptyl carbamate of (-)-eseroline. Compared to physostigmine, this compound has a long duration of action. The structures of various pseudo-irreversible AChEIs are presented in Figure 8.

Transition State Analogue Inhibitors

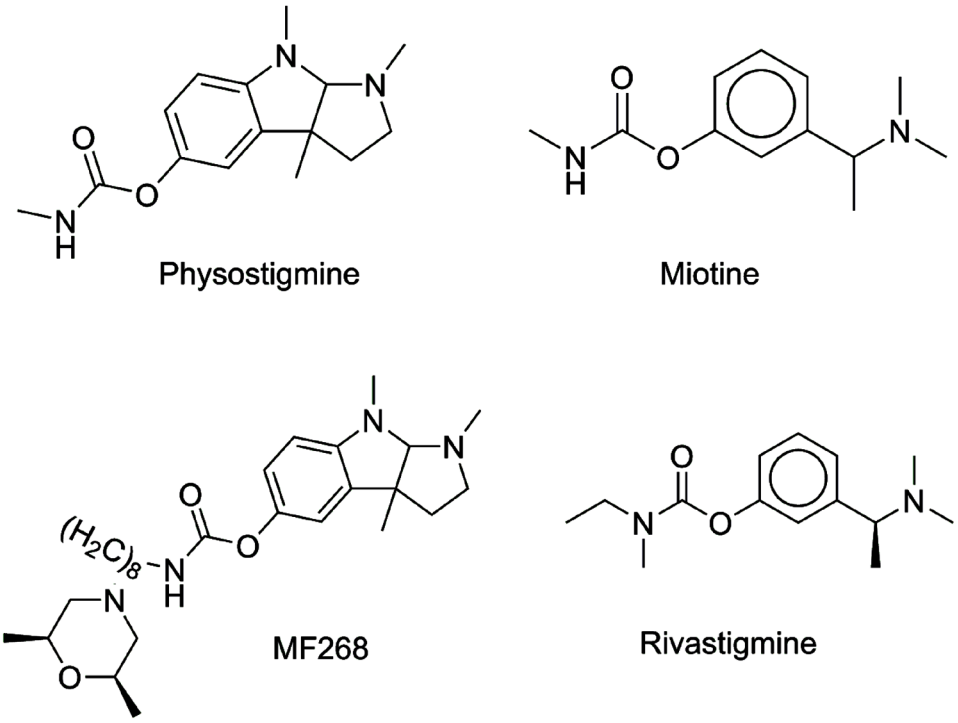
Trifluoromethylketones are effective inhibitors of this group having a reversible covalent interaction with S200 of the AChE active site forming a tetrahedral-hemiketal transition state. In fact, among AChEIs, m-(N,N,N-trimethylamino) trifluoro acetophenone is a highly potent reversible inhibitor (Figure 9). However, its ionic nature prevents its capability to cross the blood-brain-barrier (BBB). Therefore, a more lipophilic and non-ionic derivative zifrosilone (MDL-73745) can function better that works as its transition state analogue form.

Figure 7. Chemical structure of metrifonate



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 8. Chemical structure of pseudo-irreversible inhibitors of AChE

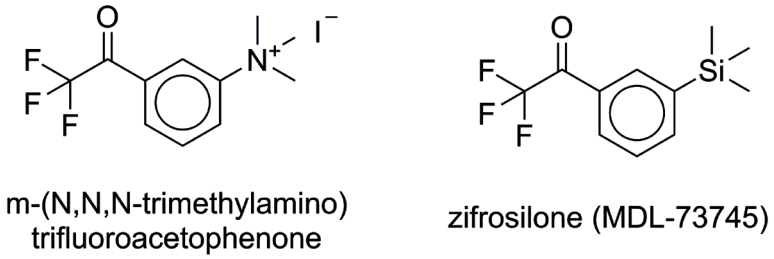


Reversible AChEIs

In contrast to the above described classes of AChEIs, reversible AChEIs binds to the active site of the enzyme and inhibits the activity of the substrate. Aminoacridines, N-benzylpiperidines and alkaloids belongs to reversible AChEIs.

- **Aminoacridines:** Tacrine (THA, Cognex®) is one of the members of this group, which was launched in 1993 as the first AChEI drug for the symptomatic treatment of AD. THA (9-amino-1,2,3,4-tetrahydroacridine) is a centrally active, reversible inhibitor of AChE that shows a moder-

Figure 9. Chemical structure of m-(N,N,N-trimethylamino) trifluoroacetophenone and zifrosilone (MDL-73745) as transition state analogue of AChEIs



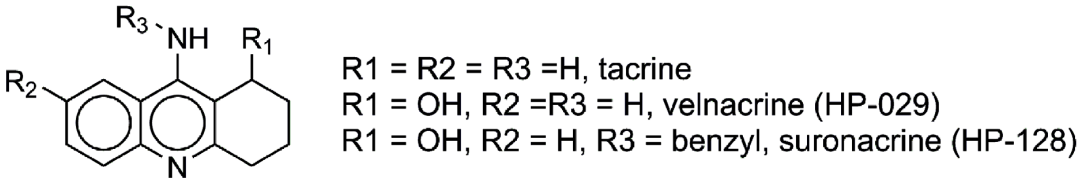
EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

ately long duration of action. Interestingly, it is more potent towards BuChE than AChE. However, it has other features including blocking sodium and potassium channels, and has a direct effect on muscarinic receptors (Schneider, 2000). Some disadvantages of these inhibitors are short half-life and induction of hepatotoxicity. However, it was the lead compound for synthesizing new derivatives, such as velnacrine and suronacrine (Figure 10), showing reduced toxicity (Ros, Aleu, Marsal, & Solsona, 2000).

- **N-Benzylpiperidines:** Donepezil (E2020) is the second drug approved by FDA for the treatment of mild to moderate AD. Donepezil, a highly-selective, reversible AChEI developed by Eisai Company in Japan, is the marker leader due to its once-daily dosing (Heydorn, 1997). It is a potent, long-acting and selective AChEI, than that of BuChE. The X-ray structure shows that the elongated donepezil molecule spans the entire length of the active-site gorge of the enzyme. It thus interacts with both the anionic subsite at the bottom of the gorge, and with the PAS, near its entrance, via aromatic stacking interactions with the conserved aromatic residues. TAK-147 (Figure 11) is another N-benzylpiperidine derivative, which has less potency than donepezil, but its effect on animals showed fewer side effects and is undergoing clinical trials. Other N-benzylpiperidine derivatives have been introduced, in which indanone moiety of donepezil has been replaced by different heterocyclic systems (Martinez, Fernandez, Castro, Conde, Rodriguez-Franco, Baños, & Badia, 2000) such as N-benzylpiperidine benzisoxazoles (Figure 11). One of the benzisoxazole derivatives is morpholino substituted molecule, which showed higher potency and selectivity than donepezil and was very effective in animal models (Villalobos, Blake, Biggers, Butler, Chapin, Chen, Ives, Jones, & Liston, 1994).
- **Alkaloids:** Galanthamine (Reminyl) is a tertiary amine alkaloid and a selective competitive AChEI. It is 50 times more effective against hAChE than BuChE at therapeutic doses. Galanthamine has been approved in several countries for the treatment of AD (Sramek, Frackiewicz, & Cutler, 2000). It enhances the response of nicotinic receptors to ACh, which causes increasing ACh release and other neurotransmitters. Furthermore, it increases the bioavailability of ACh to inhibit AChE. Galanthamine derivatives are synthesized, which include P11012 and P11149 (Figure 12) and which are 10-fold more potent and 6-fold more selective than galanthamine.

(-)-Huperzine A (Figure 13) is another alkaloid isolated from herb Huperzia serrata. Currently it is available not as a drug, but as a dietary supplement. It is a very potent, selective and long-acting AChEI with mild adverse effects include sleeping, nausea, and vomiting. Tests have shown that huperzine A also have neuroprotective properties. It has been a lead compound to design derivative such as the 10-methyl substituted that has 8-fold more potent than (-) -Huperzine (Figure 13) (Xiao, Yang, & Tang, 1999).

Figure 10. Chemical structure of tacrine, velnacrine, and suronacrine



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 11. Structure of donepezil (E2020), TAK-147 and N- benzylpiperidine morpholino-benzisoxazole

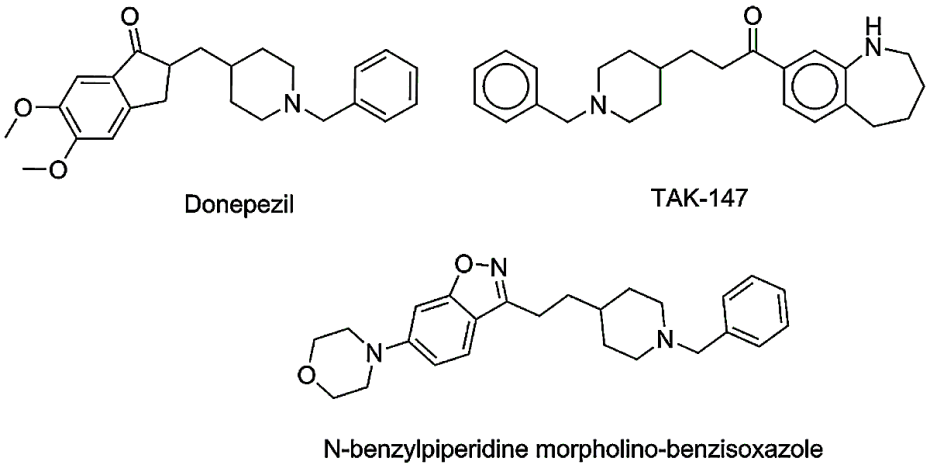


Figure 12. Chemical structure of galanthamine, P11012 and P11149

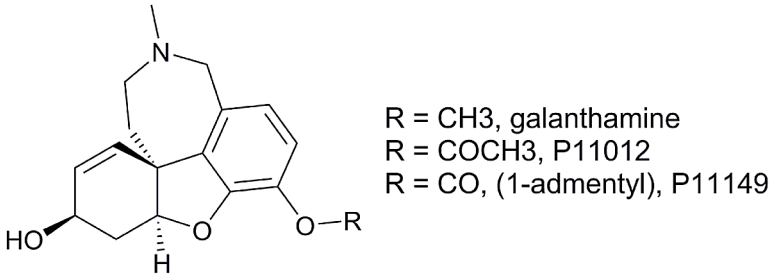
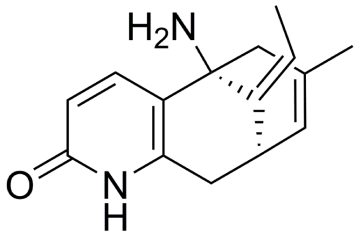


Figure 13. Chemical structure of (-) -huperzine a



Natural Non-Alkaloid

At present, attempts have been made to discover a new non-alkaloidal AChEI to avoid the side effects that have been recorded with alkaloids. Most of the other inhibitors found are terpenes (Hostettmann, Borloz, Urbain, & Marston, 2006). Synergistic effects between monoterpenes were established. The oil or extract of two species of Salvia (Salvia lavandulifolia and Salvia officinalis) has shown improvement of cognitive function in animals, and for both healthy and early stages of AD patients. The roots of Withania somnifera have been used for 4000 years in Ayurvedic medicine for the treatment of cognitive

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

decline associated with age. The triterpene withaferin A (Figure 14) has shown promising results in animal models. Among the coumarins scopoletin, isolated from Vaccinium oldhami, shows interesting activity against AChE enzyme (Houghton, Ren, & Howes, 2006; Fallarero, Oinonen, Gupta, Blom, Galkin, Mohan, & Vuorela, 2008).

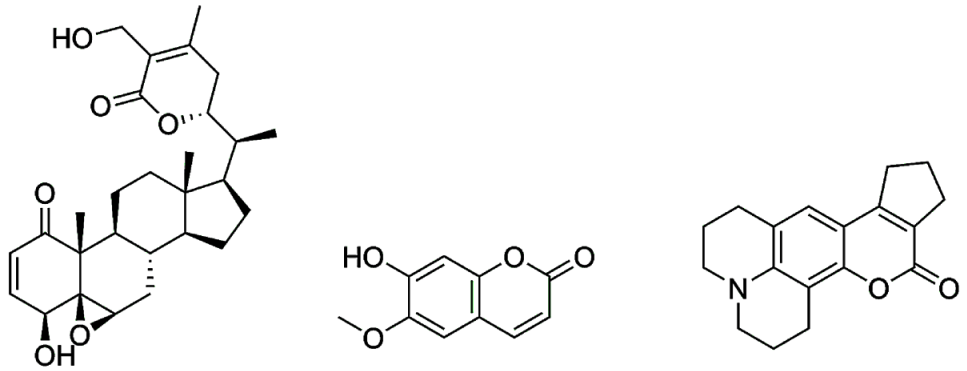
Butyrylcholinesterase (BuChE) Enzyme

Cholinergic hypothesis have been proven to be the most successful therapeutic approach adopted for symptomatic relief on AD, till today by the effective use of cholinesterase inhibitors. According to this hypothesis many of the cognitive, functional, and behavioral symptoms faced by AD patients are a result of reduced levels of cholinergic neurotransmission (Holzgrabe, Kapkova, Alptüzün, Scheiber, & Kugelmann, 2007). Hence, the involvement of cholinergic neurons causes decrease levels of ACh within synapses. AChE levels also decline, to compensate for the loss of ACh, while activity of other cholinesterase enzyme BuChE (EC 3.1.1.8) increases. A significant amount of ACh is metabolized by this enzyme as the disease progresses. Another promising approach is the development of dual inhibitors for AChE and BuChE (Fallarero, Oinonen, Gupta, Blom, & Galkin, 2008). BuChE activity seems to correlate with AChE activity in AD and a cognitive improvement could be reached (Decker, Kraus, & Heilmann, 2008). Also, many AChEIs also inhibit BuChE, because both AChE and BuChE enzymes are found in the CNS. Moreover, AChE and BuChE share 65% amino acid sequence homology even though being encoded by different genes on human chromosomes (La Du, 1994). It has been observed that in advanced AD, BuChE hydrolyses the already exhausted ACh levels, thus exacerbating the cholinergic imbalance (Greig, Utsuki, & Lahiri, 2005). Further, the low-activity BuChE in AD patients directly correlates with better cognitive function of the brain.

AN OVERVIEW OF DUAL BINDING SITE ACHEIS

The particular architecture of the enzyme AChE, in which the CS and PAS were close enough as to be simultaneously spanned by a single not too large molecule, had driven the rational design of novel classes

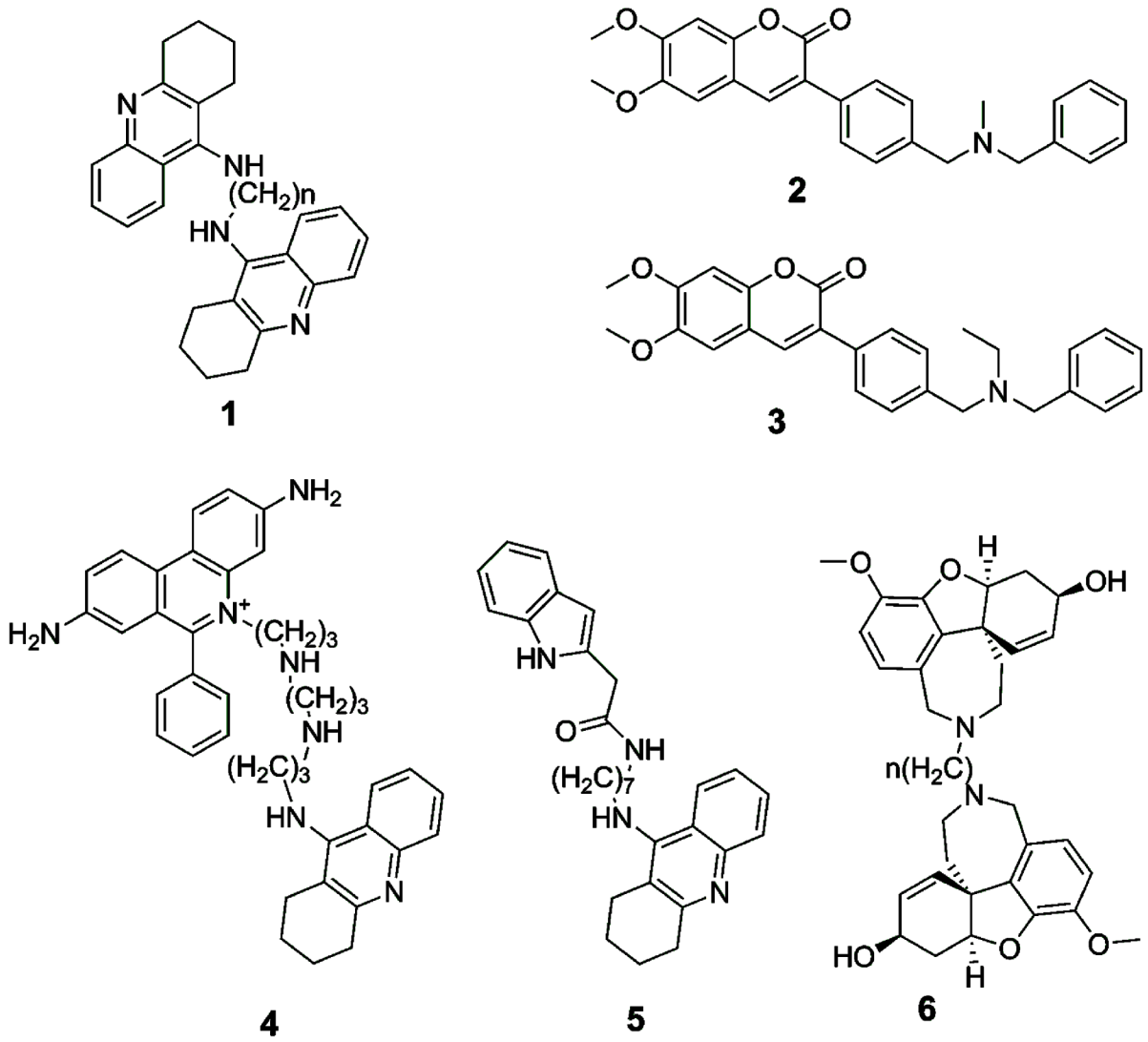
Figure 14. Chemical structure of withaferin a, scopoletin, and coumarin106



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

of a dual binding site AChEIs (Inestrosa, Alvarez, Perez, Moreno, Vicente, Linker, Casanueva, Soto, & Garrido, 1996). The sole dual binding site AChEI approved for the treatment of AD is donepezil, and at 100 μM it can inhibit 22% of the AChE-induced aggregation of Aβ (Bartolini, Bertucci, Cavrini, & Andrisano, 2003). In 1996, the first family of a dual binding site AChEIs was developed by the group of Pang and Carlier, in which the two component units are connected through a heptamethylene chain (Pang, Quiram, Jelacic, Hong, & Brimijoin, 1996). Indeed, the most potent of Pang's compounds, the so-called bis(7)- tacrine (**1**, Figure 15), turned out to be 150-fold more potent than the parent compound tacrine as an inhibitor of rat brain AChE. It is also 250-fold more selective towards AChE than BuChE. Bis(7)-tacrine has emerged as a very promising anti-Alzheimer drug candidate for its wide range of neuroprotective effects against a variety of neuronal injury models (Du, & Carlier, 2004).

Figure 15. Different dual binding site AChE inhibitors with Aβ anti-aggregating effects



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

The success of the dual binding site strategy is evidenced by the large increase in AChE inhibitory potency of these dimers or hybrids relative to the parent compounds from which they have been designed. Campiani *et al.* has reported that replacement of the central methylene group of bis(7)-tacrine by a protonatable methylamino group and capable of providing additional specific interactions with this mid-gorge recognition site leads to the dramatic increase in potency (Savini, Gaeta, Fattorusso, Catalanotti, Campiani, Chiasserini, Pellerano, Novellino, McKissic, & Saxena, 2003). Analogously, the group of Valenti and Recanatini has developed a very interesting AChEI, AP2238 (**2**, Figure 15), consisting of benzylamino and coumarin moieties as the dual binding units, respectively, connected by a *p*-phenylene linker and able to interact with the aromatic residues at the enzyme gorge. In addition to its high AChE inhibitory potency, AP2238 exhibits a significant β-amyloid anti-aggregating action (Piazzi, Rampa, Bisi, Gobbi, Belluti, Cavalli, Bartolini, Andrisano, Valenti, & Recanatini, 2003). In order to improve the moderate potency of AP2238 toward both *human* AChE (*h*AChE) and *h*AChE-induced Aβ aggregation, Piazzi *et al.* has recently reported the synthesis of a large series of AP2238 derivatives (Piazzi, Cavalli, Belluti, Bisi, Gobbi, Rizzo, Bartolini, Andrisano, Recanatini, & Rampa, 2007). The suitability of the o-methoxy-substitution in AP2238 derivatives could be analogously related to its positive effect on the basicity of the nitrogen atom of the N-methyl-substituted o-methoxybenzylamino moiety and consequently on its extent of protonation.

The derivative of AP2248 leads to a higher potency after replacing the methyl group on the nitrogen atom of the benzylamino moiety by an ethyl group AP2243 (**3**, Figure 15). Melchiorre and Bolognesi developed a new dual binding site AChEI (**4**, Figure 15), composed of a unit of tacrine as the active site interacting moiety and the phenantridinium motif of propidium as the PAS interacting moiety, connecting both units through a triamine linker (Bolognesi, Andrisano, Bartolini, Banzi, & Melchiorre, 2005). This hybridization resulted in a 20000- and 300-fold increase in *h*AChE inhibitory activity relative to the parent compounds propidium and tacrine, respectively. Neuropharma, developed a new family (**5**, Figure 15), of tacrine-based dual binding site AChEIs which are the most potent inhibitors of AChE-induced Aβ aggregation and among the most potent inhibitors of *h*AChE so far reported. Greenblatt *et al.* designed and reported the bifunctional derivatives (**6**, Figure 15) of the alkaloid galanthamine that interacted with both the active site and PAS of the AChE (Greenblatt, Guillou, Guénard, Argaman, Botti, Badet, Thal, Silman, & Sussman, 2004). Bermúdez-Lugo *et al.* reviewed several computational methods for designing AChEIs, such as docking approaches, molecular dynamics studies, quantum mechanical studies, electronic properties, hindrance effects, partition coefficients (Log P) and molecular electrostatic potentials surfaces, among other physicochemical methods that exhibit QSARs. (Bermúdez-Lugo, Rosales-Hernández, Deeb, Trujillo-Ferrara, & Correa-Basurto, (2011).

Quantitative Structure-Activity Relationship

Quantitative structure-activity relationship (QSAR) and quantitative structure-property relationship (QSPR) are based on the assumption that the structure of a compound (i.e. its geometric, steric and electronic properties) must contain features responsible for its physical, chemical, and biological properties (Hansch, Hoekman, & Gao, 1996). More than a century ago, Crum-Brown and Fraser (1868) expressed the idea that the physiological action of a substance in a certain biological system (Φ) was a function (f) of its chemical constitution C:

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Biological Activity (Φ) = f Physicochemical Properties(C)

QSAR/QSPR attempts to find out the quantitative correlation between physicochemical parameters of compounds (independent variable) and their pharmacodynamic, pharmacokinetic, toxicological properties or biological activities (dependent variables) (Yap, Li, Ji, & Chen, 2007). Determining a QSAR generally proceed as follows:

1. Define a quantitative measure of activity (e.g., the amount of ligand needed to produce an interference with the functioning of the target).
2. Express the ligand in some quantitative manner; that is, select a collection of numbers that characterize the ligand. These numbers are called 'molecular descriptors' or, 'descriptors.'
3. Determine a functional relationship between activity and the selected descriptors; that is, the search for a mathematical function, *f*, that has the property that "activity = *f* (descriptors)" to a suitably high level of accuracy.
4. Use the know activity values, molecular descriptors and determined functional relationship to predict the activity of new candidate compound.

QSAR methods are used to generalize experimental data in order to design or optimize new biologically active compounds that are more potent, less toxic, more selective, or satisfy other relevant criteria (Winkler, 2002). Once, a reliable QSAR model is created, it is possible to predict the activity of new compounds, and to learn which structural properties play an important role in the modeled biological response. It can also identify and describe important structural features of the compounds that are relevant to variations in molecular activities. The most widely used method for determining the functional relationship is the statistical technique of regression or least squares. Various multi-dimensional descriptors, especially 2D and 3D descriptors of molecular properties (particularly steric and electrostatic values) have been defined that can be correlated with the biological activity using statistical or machine learning techniques (Li, Yap, Ung, Xue, Li, Han, Lin, & Chen, 2007). An indispensable prerequisite for QSAR/QSPR model generation is the set of compounds with reliable measured bioactivity, also known as 'training set molecules'. The model is validated by various methods and final model is used to predict the activity of new compounds in the series. A number of automated computer programs are available for the fast generation of various descriptors and QSAR models (Roy, 2007). Apart from drug discovery research the QSAR is being applied in many other disciplines. In fact a QSAR model can be built to predict any type of physical property (Rücker, Meringer, & Kerber, 2004; Sutter, & Jurs, 1996) or biological activity/toxicity, of a given set of observations and molecular structures (Novak, & Rajagopal, 2002; Gupta, 2007).

HISTORY OF QSAR

Historically, at the turn of the 20th century, Meyer (1899) and Overton (1895) independently suggested that the narcotic (depressant) action of a group of organic compounds paralleled their olive oil/water partition coefficients. In following years on the physical organic front, the seminal work of Hammett

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

gave rise to the "σ−ρ" culture (Hammett, 1935) in the delineation of substituent effects on organic reactions, while Taft devised a way for separating polar, steric, and resonance effects and introducing the first steric parameter, ES (Taft Jr, 1952).

In 1962, Hansch and Fujita published their first 2D-QSAR study on the plant growth regulators and their dependency on Hammett constants and hydrophobicity (Hansch, Maloney, Fujita, & Muir, 1962). Using the octanol/water system, a whole series of partition coefficients were measured, and thus a new hydrophobic scale was introduced (Hansch, & Fujita, 1964). The contributions of Hammett and Taft together laid the basis for the development of the QSAR paradigm by Hansch and Fujita, which combined the hydrophobic constants with Hammett's electronic constants to yield the linear Hansch equation and its many extended forms.

Types of QSAR Models

QSAR is one of the basic tools of modern drug design science. It is founded on the systematic use of mathematical models from the multivariate point of view. Further, it has an increasing role in environmental sciences. QSAR models exist at the intersection of chemistry, statistics and biology (toxicology) studies. Data used in QSAR evaluations are obtained either from the literature or generated specifically for QSAR analysis and should be accurate and precise. It can consist of congeneric series of compounds or assure structural diversity even within a chemical class. This diversity has allowed the generalization of more robust QSARs, applied in an extended way. A structure– activity model is defined and limited by the nature and quality of the data used in model development and should be applied only within the model's applicability domain.

Nowadays, several multidimensional QSAR approaches (3D-QSAR, 4D-QSAR, 5D-QSAR and 6D-QSAR) are also available depending upon the type of descriptors used (Lill, 2007). It is important to note that all QSAR approaches aims to correlate structural/compounds data with the biological activity or some other properties (Kubinyi, 1997; Dudek, Arodz, & Galvez, 2006). The 3D-QSAR methodology is much more computationally complex than the 2D-QSAR approach. In general, it involves several steps to obtain different numerical descriptors of the compound structure. First, the conformation of the compound has to be determined either from experimental data or molecular mechanics and then the resulting conformations are refined by minimizing the energy (Akamatsu, 2002). Next, the conformers in dataset have to be uniformly aligned in space. Finally, the space with immersed conformer is probed computationally for various descriptors. Examples of 3D-QSAR are Comparative Molecular Field Analysis (CoMFA) (Cramer Iii, Patterson, & Bunce, 1998), Comparative Molecular Similarity Indices (CoMSIA) (Klebe, Abraham, & Mietzner, 1994) and Self Organizing Molecular Field Analysis (SOMFA).

Some methods independent of the compound alignment have also been developed. Alignment-independent 3D-QSAR descriptors are 3D descriptors that are invariant to molecular rotation and translation in space. Thus, no superposition of compounds is required. The Comparative Molecular Moment Analysis (CoMMA) (Silverman, & Daniel, 1996) uses second-order moments (moments relate to the center of the mass and center of the dipole) of the mass distribution and charge distributions. The VolSurf approach is based on probing the grid around the molecule with specific probes, e.g. the hydrophobic interactions or hydrogen bond acceptor or donor groups (Crivori, Cruciani, Carrupt, & Testa, 2000). 3D-QSAR models are widely applicable, but often suitable for any series of congeneric compounds.

The broad family of chemical descriptor(s) includes empirical, quantum chemical or non-empirical parameters used in the 2D-QSAR approach share a common property of being independent from the 3D

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

orientation of the compound. Empirical descriptors may be measured or estimated and include physicochemical properties. Non-empirical descriptors can be based on individual atoms, substituents, or the whole molecule, they are typical structural features. They can be based on topology or graph theory and, as such, they are developed both from the knowledge of 2D structure, or from the 3D structural conformations of a compound. Some of the different types of descriptors and their description are presented in Table 1.

Different software calculates wide sets of different theoretical descriptors, from SMILES, 2D-graphs to 3D- coordinates etc. Some of the software includes: ADAPT (Stuper, & Jurs, 1976), OASIS (Mekenyan, & Bonchev, 1986), CODESSA (Katritzky, Petrukhin, Yang, & Karelson, 2005), MolConZ, and the DRAGON (Todeschini, Consonni, Mauri, & Pavan, 2003). Thus, descriptor selection must be performed by applying mathematical approaches with the final and crucial aim to maximize, as an optimization parameter, the predictive power of the QSAR model, as the real utility of any model is considered its predictivity. For the potency modelling, the most widely used mathematical technique is multiple regression analysis (MRA). Regression analysis is a simple approach that leads to a result that is easy to understand. For the modelling of categories, a wide range of classification methods exists, including: discriminant analysis (DA), SIMCA (Soft Independent Modeling of Class Analogy), k-NN (k-Nearest Neighbours), CART (Classification and Regression Tree), Artificial Neural Network (ANN), Support Vector Machine (SVM), etc. In these techniques, the term "quantitative" is referring to the numerical value of the variables (descriptors) necessary to classify the chemicals in the qualitative classes. Nowadays, genetic function algorithm (GFA) has gained great popularity in QSAR research. GFA and genetic PLS (G/PLS) methods developed by Rogers and Hopfinger (1994), employed to select the relevant descriptors.

Linear and Non-Linear Statistical Regression Techniques

- **GFA:** GFA is genetics based linear method of variable selection that combines Holland's genetic algorithm (GA) with Friedman's (1991) multivariate adaptive regression splines (MARS). It works in the following way: First, a predefined number of equations (set at 100 by default) are generated randomly, then pairs of ''parent'' equations are chosen randomly from this set of 100 equations and the ''crossover'' operations are performed at random. After some preliminary observations on the initial runs, the number of GFA crossovers was set to 5000 for the present study

Table 1. Types and description of various descriptors employed for obtaining QSAR models

Type	Descriptors Description
Electronic	Sum of partial charges, sum of formal charges, dipole moment, HOMO, LUMO
Spatial	Radius of gyration, shadow indices, area, density, PMI, Vm
Structural	Molecular weight, number of rotatable bonds, number of hydrogen bond acceptor, number of hydrogen bond donors
Thermodynamics	Log of partition coefficient, log of partition coefficient atom-type value, desolvation free energy of octanol, heat of formation, molar refractivity
Topological	Wiener index, Zagreb index, Kier and Hall molecular connectivity index, Balaban index, Hosoya index
E-State Indices	Electrotopological-state indices

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

in order to obtain a reasonable convergence. The goodness of each progeny equation is assessed by Friedman's lack of fit (LOF) scoring, assigned by GFA, which resists over the fitting and estimates the appropriate number of variables in the equation. The LOF is given by the equation,

$$LOF = LSE / \left\{1 - \left(c + dp\right) / m\right\}^2 \quad (1)$$

where LSE is the least-squares error, c is the number of basis functions in the model, d is the smoothing parameter, p is the number of descriptors, and m is the number of observations in the training set. The smoothing parameter that controls the scoring bias between the equations of different size was set at a default value of 1.0. A new term was added to the offspring models with a probability of 50%. Only the linear equation terms were used for model building. The best equation out of 100 equations was taken based on the statistical parameters such as regression coefficient, adjusted regression coefficient, regression coefficient cross validation and F-test values.

- **G/PLS:** G/PLS methodology combines GFA, and is used for the selection of descriptors to be included in a model. In the partial least squares (PLS) regression fitting technique (Roy, Roy, 2008) that weights the relative contribution of each descriptor in the final model. PLS in particular produces statistically significant solutions by identifying a linear combination of the original physico-chemical descriptors that best correlates with the biological response (Dunn Iii, Scott, & Glen, 1989). The main advantage of the G/PLS method is that it constructs QSAR equations from which most variables have been eliminated; thus, PLS avoids over fitting. Application of G/PLS thus allows the construction of larger QSAR equations, while still avoiding over fitting and eliminating most variables.
- **SVM:** Support vector machine (SVM), first developed by Vapnik and Cortes (1995) is a novel machine-learning non-linear method for classification problems. The main advantage of SVM is that it was built on the structure risk minimization principle, which has been shown to be superior to the traditional empirical risk minimization principle. After the introduction of ε-insensitive loss function, SVM has been able to solve non-linear regression estimation and has demonstrated much success in QSAR studies (Yuan, Zhang, & Luo, 2009). In brief, the correlation between the structure of the training set and its biological activity, can be represented as $\mathbf{Yi = f\ (x_i)}$. Simply, we can represent $\mathbf{f\ (x_i)}$ as a linear function of

$$y_i = w_i x_i + b \quad (2)$$

where, w_i is the coefficient vector of the linear function and *b* corresponds to its coefficient.

In SVM, the basic idea is to map the data x into a higher-dimensional feature space **F** through a non-linear mapping **ϕ** by using kernel function and then performed linear regression in this space. In the high dimensional feature space, SVM approximates the set of data with a linear function:

$$Y = f\left(x\right) = \sum\nolimits_{i=1}^{n} w_i \dot{\phi}\left(x_i\right) + b \quad (3)$$

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

where $\dot{\phi}\left(x_i\right)$ is the high-dimensional feature space after kernel transformation, which is nonlinearly mapped from the input space x, while w_i and b are coefficients. A radial basis function (RBF) kernel is frequently used kernel function in QSAR studies, out of various available kernel functions for nonlinear transformation of the input data. The most important aim of SVM is minimization of the regularized risk function, by which the coefficients w and b are estimated. The purpose of this risk function is to find a function that has at most ε prediction errors in all the training data points and at the same time is as flat as possible. The regularized risk function is defined as

$$R(C) = C\frac{1}{N}\sum_{i=1}^{N} L\varepsilon(y_i, f(x_i, w)) + \frac{1}{2}\left\|\mathrm{w}\right\|^2 \tag{4}$$

$$L\varepsilon\left(y, f\left(x,w\right)\right) = \begin{matrix} \left\{\left|y - f\left(x,w\right)\right|\right\} - \varepsilon, & for\ \left|y - f\left(x,w\right)\right| \geq \varepsilon \\ 0 & otherwise \end{matrix}$$

The first term of the equation $C\frac{1}{N}\sum_{i=1}^{N} L\varepsilon(y_i, f(x_i, w))$ is the empirical error (risk), which is estimated by the ε insensitive loss function $L\varepsilon\left(y, f\left(x,w\right)\right)$. The term, $\frac{1}{2}\left\|w\right\|^2$ is known as the regularized term and is used as a measurement of function flatness. The parameter ε is known as the tube size, and it is equivalent to the approximation accuracy placed on the training data points, and C is the regularization constant determining the trade-off between the training error and the regularized term. The minimization of regularized risk function is constrained optimization problem which can be reformulated into dual problem formalism by using Lagrange multipliers. The performance of SVM regression is based on the kernel selection, and the optimized value of C and ε. Parameter tuning must be performed because, which values of C and ε will work the best for a particular data set is not known previously.

- **ANN:** Artificial neural network (ANN) method is an interconnected feed forward network, which is motivated by the way biological nervous systems, *viz.* the brain process information, and is a popular tool in function learning (Gasteiger & Zupan, 1993). It has the ability to learn complicated nonlinear function (dataset) with good efficacy. The theory of the ANN and its application in QSAR/QSPR studies were described extensively in many reviews (Zupan, 1994). An ANN technique in general has three interconnected layers: input layer, a hidden layer, and an output layer. The input layer neurons receive the data from molecular descriptors input files, which is subsequently passed on to the nodes of the hidden layer for further processing. The signals are then relayed onto the output layer, which sends information directly to the outside world, to a secondary computer process or to other devices such as a mechanical control system. The connections between the nodes of each layer are assigned by randomized weight value. Supervised learning technique was used in back-propagation ANN, the network is trained by minimizing the squared error of the network's output. This is achieved by back propagating the desired output to the transfer function (sigmoid function) for adjustment of weight by using a gradient descent algorithm.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

QUALITY OF FIT AND QSAR MODEL VALIDATION METHODS

The success of a QSAR model is to measure the quality of fit on the available training data set. The most common statistical qualities of the equations were measured by the parameters such as explained variance (R^2_a), squared correlation coefficient (R^2) and variance ratio (F-test) at specified degrees of freedom (Wold, & Eriksson, 1995). While for G/PLS equations, least-squares error (LSE) was taken as a statistical measure, lack-of-fit (LOF) was noted for the GFA derived equations. The expression for R^2, R^2_a and the F-test are defined as mentioned below:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y}_{Training})^2} \tag{5}$$

$$R_a^2 = \frac{(N-1)R^2 - d - 1}{N - d - 1} \tag{6}$$

$$F = \frac{(R^2 / d)}{((1 - R^2) / (N - d - 1))}. \tag{7}$$

In the above Equations (5-7), $\hat{Y}$, i =1,…,n are the values calculated by the QSAR model for the dependent variable corresponding to object 'i', when this compound has been included in the training data set. Whereas, $\bar{Y}_{Training}$ is the averaged value of the dependent variable for the training set, *N* is the number of compounds in the training set, and *d* is the number of descriptors in the QSAR equation.

As suggested by Tropsha, Gramatica, & Gombar, (2003) the predictive ability of a 2D-QSAR model should be tested on an external set (test set) of data that has not been taken into account during the process of developing the model. Hence, the best models were used to predict the enzyme inhibition potency value of the test set compounds. The prediction qualities of the models were judged by statistical parameters like predictive correlation coefficient R^2_{pred}, squared correlation coefficient between actual and predicted values, with (r^2) and without (r_0^2) intercept of the test set compounds. The predictive value is calculated as follows:

$$R^2_{pred} = 1 - \frac{\sum\left(Y_{pred(\text{Test})} - Y_{(Test)}\right)^2}{\sum\left(Y_{(Test)} - \bar{Y}_{(Training)}\right)^2}. \tag{8}$$

In the above equation, $\boldsymbol{Y}_{\boldsymbol{pred\ (Test)}}$ and $\boldsymbol{Y}_{\boldsymbol{(Test)}}$ indicate predicted and actual activity values, respectively, of the test set compounds, and $\bar{Y}_{(Training)}$ indicates the mean activity value of the training set compounds. It was previously shown that the use of R^2_{pred} and r^2 might not be sufficient to indicate the external

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

validation characteristics. Thus, an additional parameter r_m^2 defined as $r_m^2 = r^2 * \left(1 - \sqrt{r^2 - r_0^2}\right)$ which penalizes a model for large differences between the actual and the predicted values was also calculated (Roy, 2007).

QSAR Model Cross-Validation

Cross-validation is a popular technique which is used to explore the predictive ability of the QSAR models. These models were rigorously evaluated using leave one out (LOO) or leave many out (LMO) approach, Y-randomization test and test set predictions. In the LOO or LMO approach, one compound or small group of compounds at a time is removed from the training set, the model is recalibrated on the reduced data set, and the biological response value is predicted for the removed compound using the modified model. The procedure is repeated for each compound in the training data set. The result from the cross-validation procedure is validated through R_{LOO}^2, predicted residual sum of squares (PRESS), and root mean square error (RMSE), which is used as a criterion of both robustness and predictive ability of the QSAR models (Baumann, 2003). The maximum number of models produced by the LOO procedure is equal to the number of available examples 'n'. The standard deviation of the prediction (S_{PRESS}) statistics can also be calculated. Cross-validated correlation coefficient R_{LOO}^2, and S_{PRESS}, is calculated according to the formula defined below:

$$R_{LOO}^2 = 1 - \frac{\sum_{i=1}^{n} (Y_i - \tilde{Y}_{i\,LOO})^2}{\sum_{i=1}^{n} (Y_i - \bar{Y}_{Training})^2} \tag{9}$$

$$S_{PRESS} = \sqrt{\frac{PRESS}{(N - d - 1)}} \tag{10}$$

$$RMSE = \sqrt{\frac{PRESS}{N}}\,. \tag{11}$$

In the above Equations (9-11), $\bar{Y}_{(Training)}$ is the mean average activity values of the training set, while Y_i, $\tilde{Y}_{i\,LOO}$ i =1,...., n are the measured and predicted values of the dependent variable over the available training set. Often, a high R_{LOO}^2 value ($R_{LOO}^2 > 0.5$) is considered as a proof of high predictive ability of the QSAR model (Todeschini, Consonni, Mauri, & Pavan, 2004).

Furthermore, the developed QSAR models were subjected to a randomization test for validation purpose. The Y-randomization test is used to assess whether a model is obtained by a chance correlation (Rücker, Rücker, & Meringer, 2007; Golbraikh, & Tropsha, 2002). The dependent variable vector is randomly shuffled and a new QSAR model is developed using the original independent variable matrix.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

The new QSAR models are expected to have low R^2 values after several iterations. If the opposite happens, then an acceptable QSAR model cannot be obtained for the specific modeling method and data. The randomization experiment was further validated by computing the ${}^{c}R_{p}^{2}$ metric proposed by Mitra, Saha, and Roy (2010) ${}^{c}R_{p}^{2} = R\sqrt{\left(R^2 - R_{r}^{2}\right)}$. Finally, the predictive accuracy of the developed QSAR models was evaluated more rigorously by test set compounds.

Thus, among the available *in silico* techniques QSAR is one of the most important areas in chemoinformatics, and its advances have widened the scope of rational drug design and the search for the mechanism of drug action (Gupta, 2007). Over the past few decades, a variety of QSAR models for AChEIs have been developed by different research groups to improve selectivity and ADMET properties (Garg, Gandhi, & Mohan, 2008). The developed models should be able to identify and describe important physico-chemical features of the compounds that underpin variations in molecular activity for designing potent and safe AChEIs.

COMPUTATIONAL APPROACHES ON DRUG DESIGN

In the past few decades, drug discovery efforts have shifted toward the design of more potent and selective AChEIs with fewer side effects. The interest in the discovery of novel potent AChEIs is expected to continue in the future, as current AChEIs lack perfection. In current scenario, direct- and indirect-computer-aided drug design (CADD) techniques are of utmost importance in terms of being time-saving and cost-effective techniques. One of the rationale starting points for the design of novel scaffolds is the development of a ligand-based predictive model to derive different structural features essential for receptor binding, using ligand-based molecular modeling tools (Ooms, 2000, John, 2013). CADD is also known as computer-assisted molecular design (CAMD) has undergone a paradigm shift in the past few years. Modeling and informatics have become indispensable components of rational drug design (RDD). Previously, chemical analysis through molecular modeling has been very prominent in CADD. However, currently modeling and informatics are contributing in tandem toward CADD. Although, CADD is a topic of medicinal chemistry and before venturing into this exercise one must employ computational chemistry methods to understand the properties of chemical species, on the one hand, and employ computational biology techniques to understand the properties of biomolecules on the other.

Information technology is playing a major role in decision making in pharmaceutical sciences. Storage, retrieval, and analysis of data of chemicals/biochemicals of therapeutic interest are major components of pharmacoinformatics. We know that the new drug discovery is a comprehensive, expensive and time consuming process. It is estimated that a drug from concept to market would take around 12-14 years and cost more than US$1.3 billion on an average (Malik, 2008). Traditional way of new drug discovery involves the identification of a lead compound (natural or synthetic) through random screening for some desired biological activity and building a library around the scaffold to find the most potent and efficacious analog. This is an iterative process and generates the structure-activity relationship (SAR) that may further help to design more molecules until the desired need is met. In recent years, RDD has emerged as a widely used approach for the drug design and discovery in order to reduce the time and expenses (Gomeni, Bani, D'Angeli, Corsi, & Bye, 2001). CADD or *in silico* drug design is one of the RDD technologies that is now widely used in industrial and academic drug discovery and development programs. It is also estimated that computer modeling and simulations account for about 10% of pharma-

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

ceutical R&D expenditure, and that they give rise to 20% by 2016 (Kapetanovic, 2008). In most current applications of CADD, attempts are made to find a ligand that will interact favorably with a receptor that represents the target site. Binding of ligand to the receptor may include hydrophobic, electrostatic, and hydrogen-bonding interactions. In addition, solvation energies of the ligand and receptor site also are important because partial to complete desolvation must occur prior to binding (Wang, Chiu, & He, 2005). This approach to CADD optimizes the fit of a ligand in a receptor site. However, optimum fit in a target site does not guarantee that the desired activity of the drug will be enhanced or that undesired side effects will be diminished. CADD is dependent upon the amount of information that is available about the ligand and receptor. Ideally, one would have 3D structural information for the receptor and the ligand-receptor complex either from X-ray diffraction or NMR but in the opposite extreme, one may have no experimental data to assist in building models of the ligand and receptor, in which case computational methods must be applied without the constraints that the experimental data would provide.

CADD approaches can be broadly divided into two categories:

1. Ligand based drug design (LBDD).
2. Structure based drug design (SBDD).

LBDD methods are applied when the structure of the macromolecule - target is unknown, and there is unable to design its reliable model. The main reason behind is many important receptors are membrane-bound proteins, which are notoriously difficult to crystallize. So, in that situation, a lead compound or active ligand must be found, and then the structure of the ligand guides the drug design process. These methods are based on analysis of sets of ligands with known biological activity (Taft, da Silva, & da Silva, 2008). They include: design of pharmacophore models, which represents a set of points in space with the certain properties and distances between them, which define binding on the given group of ligands with a target. Furthermore, quantitative structure-activity relationship ("classic" QSAR) is another effective LBDD method for the development of close analogues of known compounds and its modification. The 3D-QSAR methods are somewhat capable of predicting adequately pharmacological activity of compounds from different chemical classes. Where there is sufficient information, these methods can be used in conjunction to increase the accuracy of simulation and enhance the drug design process (Hillisch, Pineda, & Hilgenfeld, 2004).

SBDD is one of the several methods in the rational drug design and pharmaceutical research. The main aim in SBDD is to identify chemical compounds or peptides that bind strongly to key regions of biologically relevant molecules, e.g. enzymes or receptors, for which three-dimensional structures are known (Song, Lim, & Tong, 2009). The 3D structure of protein targets is most often derived from X-ray crystallography or nuclear magnetic resonance (NMR) techniques or even homology modeling. Designed chemical compounds should be able to inhibit or stimulate the biological activity of their target molecules (Anderson, 2003). Pang *et al.* has performed docking studies for a series of benzylamino AChEIs, which was mainly focused on obtaining correct binding modes for the bound ligands (Pang, & Kozikowski, 1994). Further, Pang and Kozikowski have performed docking simulations for huperzine A and E2020 (aricept) into a series of 69 AChE conformers chosen from molecular dynamics (MD) studies so as to approximate the receptor flexibility space available to an inhibiting ligand (Pang, & Kozikowski, 1994). Correa-Basurto *et al.* has performed molecular docking studies and density functional theory (DFT) of 88 N-aryl derivatives as AChEIs (Correa-Basurto, Flores-Sandoval, Marin-Cruz, Rojo-Dominguez, Espinoza-Fonseca, & Trujillo-Ferrara, 2007). Vasilyev *et al.* published a very effective QM/MM study

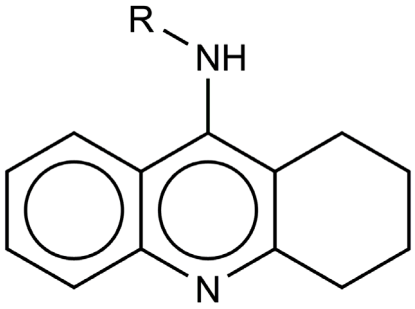
EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

by utilizing semiempirical PM3 theory and the OPLS force field, demonstrating the strong influence of hydrogen bonding within the catalytic triad on the configuration of the bare enzyme, as well as on the acylation and phosphorylation reactions (Vasilyev, 1994). Fuxreiter and Warshel studied AChE acylation through calculation of activation free energies using the empirical valence bond (EVB) and an all-atom free energy perturbation (FEP) approach, in conjunction with the PDLD/S-LRA method (Fuxreiter, & Warshel, 1998). More explicit quantum detail on the acylation reaction was provided by Zhang *et al.* who used a pseudobond *ab initio* QM/MM approach with HF/3-21G for the quantum layer and the AMBER95 forcefield for the remainder of the enzyme (Zhang, Kua, & McCammon, 2002). Bencsura *et al.* carried out quantum work on reactions within the active site related to enzyme phosphorylation (Bencsura, Enyedy, & Kovach, 1996). Further, quantum mechanical and QM/MM exploration of the phosphorylation reaction was performed by Hurley *et al* (Hurley, Wright, Lushington, & White, 2003). Recently, Zhou e*t al.* determined the catalytic reaction mechanism of AChE by Born-Oppenheimer *ab initio* QM/MM molecular dynamics simulations (Zhou, Wang, & Zhang, 2010).

Classical QSAR analysis has been proven reasonably effective to explain variations in AChE inhibitory activity and assessing, predicting AChE inhibition of new molecules. QSAR studies explored the function of various ligand substituent groups *via* simple charge, physicochemical, structural, and shape-based descriptors, achieving reasonable reproduction of experimental *in vitro* AChE inhibition results. Numerous descriptors are available to describe these physicochemical properties such as the steric, hydrophobic and electronic features of molecule which have been shown to be important for enzyme-ligand interactions. Over the past few decades, a variety of QSAR models for AChEIs have been developed by different research groups. Here we review the developed QSAR models and qualitative SAR data for different chemical series of AChEIs. AChEIs from various chemical classes had different interactions with the AChE enzyme, which affect the biological response and structure-activity relationships (SAR). QSAR can be employed to guide the designing of further potent AChEIs.

Tacrine (9-amino-1,2,3,4-tetrahydroacridine, THA) as the first anti-Alzheimer's drug approved by the FDA in 1993, which has beneficial effects on cognition in patients with AD (Knapp, Knopman, Solomon, Pendlebury, Davis, & Gracon, 1994). Tacrine was described as a reversible inhibitor of AChE and BuChE in 1961. However, soon afterwards it was found that it suffers from serious side effects such as slow pharmacokinetics and hepatotoxicity. However, the structure seems to be very attractive to introduce many substitutions on Tacrine, for improvement of its activity and removal of toxicity. Very few SAR studies have been reported for Tacrine derivatives as shown in Figure 16. QSAR model of amino group

Figure 16. N-monosubstituted 1,2,3,4-tetrahydro-9- aminoacridines
(Recanatini, et al., 1997).



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

substituted Tacrine derivatives have been generated. The bilinear correlation was established between the inhibition constant K_i with structural length parameter L (L = Verloop length parameter for R substituent) (Recanatini, Cavalli, & Hansch, 1997).

R = H, Me, $(CH_2)nCH_3$ n=1-5, $(CH_2)nR'$ [n=1-2 R'=Ph,NH_2,$CH{=}CH_2$]

$$\text{Log } 1/Ki = -0.296(0.136)\,L + 0.569(0.220)\log(\beta \times 10^{L} + 1) + 7.921(0.539) \quad (1)$$

$$n = 11,\ r^2 = 0.844,\ q^2 = 0.842,\ s = 0.142,\ F_{3,7} = 12.59,\ L_{opt} = 4.903,\ \log \beta = -4.867$$

where n is the number of molecules, r^2 is the squared correlation coefficient, q^2 is the cross-validated correlation coefficient, s is the standard deviation, F is the variance ratio. This a typical bilinear Equation (1) leads to L_{opt} (optimized length parameter), which corresponds to a minimum of activity rather than maximum in activity. However, the QSAR model suggests that compounds with extreme length parameter L values showed better activity.

Another series of Tacrine analogues with hydroxyl group at 6th position along with amino group substitution as shown in Figure 17 were synthesized (Shutske *et al.* 1989).

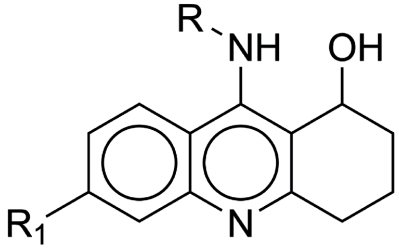
R = H, CH3, C3H7, CH2C6H4R' [R' =Me, F,CF3 etc] and R_1= H, Cl, F, OCH3, CF3

$$\text{Log } 1/IC_{50} = -0.560(0.132)\,L + 0.877(0.263)\log(\beta \times 10^{L} + 1 + 1.509(0.386)\text{I-Cl} + 7.141(0.528) \quad (2)$$

$$n = 30,\ r^2 = 0.861,\ q^2 = 0.857,\ s = 0.305,\ F_{4,25} = 38.83,\ L_{opt} = 5.116,\ \log \beta = -4.869.$$

Similar to previous QSAR model for tacrine analogues (Equation 1), this QSAR equation also possess an inverse bilinear term in L. With the L value about 5 Å, minimum AChE activity is observed. The indicator descriptor I-Cl has a value of 1 when a 6-Cl substituent is present and 0 when 6-Cl is not present, as seen from Equation (2). The substitution of –Cl group at 6th position/ R1 position is favorable for higher AChE inhibitory activity.

Figure 17. 9-amino-1,2,3,4-tetrahydroacridin-1-ol analogues (Shutske et al. 1989).



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

The earliest application of comparative molecular field analysis (CoMFA) methodology to the AChE inhibition problem, appears to be the work of Cho *et al.* who employed templates obtained from crystal structures of AChE inhibited by edrophonium, physostigmine and 9-amino-, 2,3,4-tetrahydroacridine (THA) to align a set of 60 structurally diverse known inhibitors (Cho, Garsia, Bier, & Tropsha, 1996).

On the other hand, Tong *et al.* developed a CoMFA model for a series of 1-benzyl-4[2-(N-benzoyl-amino) ethyl] piperidine derivatives and N-benzylpiperidine benzisoxazoles, finding a strong correlation between the inhibitory activity of those N-benzylpiperidines and two separate alignment considering both steric and electronic aspects of the ligand set (Tong, Collantes, Chen, & Welsh, 1996). They also included neutral and N-piperidine-protonated species for developing 3D-QSAR model. The developed model shows a strong correlation between the AChE inhibitory activity of N-benzylpiperidines and its steric and electronic factors.

Also, Hasegawa and coworkers applied a novel genetic algorithm (GA) based variable selection approach in CoMFA. They called this approach as GA-based region selection (GARGS) scheme to refine the CoMFA model. GARGS approach was tested on the polychlorinated dibenzofurans (PCDF) dataset to obtain a CoMFA model with improved predictivity (Hasegawa, Kimura, & Funatsu, 1999). A comprehensive SAR picture for 9-amino-1,2,3,4-tetrahydroacridine-based AChEIs a tacrine derivative designed on the basis of both the classical and the 3D QSAR equations was synthesized and kinetically evaluated, to test the predictive ability of the 2D and 3D-QSAR models (Recanatini, Cavalli, Belluti, Piazzi, Rampa, Bisi, Gobbi, Valenti, Andrisano, Bartolini, Cavrini, 2000).

Sulea *et al.* derived a structure-activity relationship as a function of an ensemble of different ligand conformers (Sulea, Kurunczi, Oprea, & Simon, 1998). The resulting topological pharmacophore maps corresponded well with the spatial constraints apparent in crystallographic receptor structures. Recently, Bernard and coworkers used a hybrid method to estimate the binding affinity of 82 N-benzylpiperidine derivatives by flexible docking them into the mouse AChE active site and establishing a 3D-QSAR model based on this ensemble via CoMFA analysis (Bernard, Kireev, Chrétien, Fortier, & Coppet, 1999). Sippl *et al.* also used a docking strategy for aligning ligand structures as a precursor to the CoMFA model generation (Sippl, Contreras, Parrot, Rival, & Wermuth, 2001). Chaudhary *et al.* have developed 3D-QSAR models using both CoMFA and CoMSIA models through pharmacophore-based alignment, were in good agreement with each other and demonstrated significant superiority over MCS-based alignment in terms of leave one- out (LOO) cross-validated q^2 values of 0.573 and 0.723 and the r^2 values of 0.972 and 0.950, respectively (Chaudhary, Roy, & Saxena, 2009). Developed model shows the importance of steric, electrostatic, and hydrophobicity for AChE inhibition activity. While, hydrophobic factor plays a major contribution to the AChE inhibitory activity, which is in strong agreement with the fact that the AChE is having a wide active site gorge (~20 Å) occupied by a large number of hydrophobic amino acid residues. Further, comparative 2D-QSAR analysis of three different classes of AChEIs namely, physostigmine analogues, 1,2,3,4-tetrahydroacridines and benzylamines was performed by Recanatini *et al.* (1997). Akula *et al.* developed 3D-QSAR model using CoMFA based on molecular docking on a series of 19 *bis*-tacrine compounds (Akula, Lecanu, Greeson, & Papadopoulos, 2006). In this study, the author has predicted the inhibitory activities against AChE using 3D-QSAR model and molecular docking technique with $q^2 = 0.714$, No. of components= 2, $r^2 = 0.97$, SEE = 0.082 and F= 328.4. The developed 3D-QSAR model suggests that mono-chloro heterodimeric substituents better occupy and fit in the binding sites.

In another study, Jung *et al.* has used a variable selection method to derive QSAR models for tacrine derivatives (Jung, Tak, Lee, & Jung, 2007). Authors have used variable selections of stepwise multiple

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

linear regression (MLR), genetic algorithm (GA)-MLR, and simulated annealing (SA)-MLR approaches to develop the QSAR of tacrine derivatives against AChE inhibitory activity.

The best QSAR equation was developed from simulated annealing (SA) MLR, which suggest that the hydrophobicity and rigidity of compounds increased inhibitory activity of tacrine derivatives. The QSAR models have greater explanatory capability and better prediction, with a smaller standard error than other methods. The resulting models describing the significant roles of hydrophobic and electrostatic interaction on increasing AChE inhibitory activity, while hydrophilic and topological feature of molecules was shown to decrease AChE inhibitory activity. Geldenhuys *et al.* has performed 3D-QSAR study on bis-aza aromatic quaternary ammonium analogs at the blood–brain barrier choline transporter (Geldenhuys, Lockman, Nguyen, Van der Schyf, Crooks, Dwoskin, & Allen, 2005). Choudhary *et al.* developed CoMFA model on some natural compounds acting as AChEIs (Choudhary, & Rode, 2003). Shen *et al.* performed CoMFA and CoMSIA analysis along with a docking study to explore the binding mode of 2-substituted 1-indanone derivatives with AChE (Shen, Liu, & Tang, 2007). Both CoMFA and CoMSIA models were developed successfully with a good correlation coefficient (r^2 =0.9) and cross-validated coefficient (q^2 = 0.7). From this study authors have suggested that the benzene ring of donepezil could form a π-π interaction with the side chain of Trp84. Also a more hydrophobic and bulky group with a highly positive charge to replace the small protonated nitrogen moiety could increase the binding affinity through hydrophobic, van der Waals, and cation–π forces with Trp84, Tyr330, and His440 of AChE respectively.

Martin-Santamaria *et al.* has used the comparative binding energy (COMBINE) methodology to identify the key residues that modulate the inhibitory potencies of three structurally different classes of AChEIs (tacrines, huprines, and dihydroquinazolines) targeting the catalytic active site of this enzyme (Martin-Santamaria, Munoz-Muriedas, Luque, & Gago, 2004). In another study, Roy *et al.* performed 3D-QSAR analysis on 78 molecules (52 and 26 molecules in training and test set, respectively) diverse carbamates analogs using CoMFA and CoMSIA methods (Roy, Dixit & Saxena, 2008). The highly predictive models (CoMFA $q^2 = 0.733$, $r^2 = 0.967$, predictive $r^2 = 0.732$, CoMSIA $q^2 = 0.641$, $r^2 = 0.936$, predictive $r^2 = 0.812$), well explained the variance in binding affinities both for the training and the test set compounds. In this study, the author suggests that steric, electrostatic and hydrophobic interactions play an important role in describing the variation in binding affinity. Specially, the carbamoyl nitrogen should be more electropositive; substitutions on this nitrogen should have the high steric bulk and hydrophobicity, while the amino nitrogen should be electronegative in order to have better activity.

Asadabadi *et al.* used a combinatorial feature selection method to describe the QSAR of dual site inhibitors of AChE that involved constitutional, topological, Galvez topological charge indices, empirical, molecular walk counts, atom-centered fragments, 2D autocorrelations, BCUT properties and functional group descriptors(Asadabadi, Abdolmaleki, Barkooie, Jahandideh, & Rezaei, 2009). In 2008 Saracoglu *et al.* established the structure-activity relationships (SAR) through the Electron- Topological Method (ETM) for a class of AChE inhibitors related to tacrine (9- amino-1,2,3,4-tetrahydroacridine) and 11 H-Indeno-[1,2-b]- quinolin-10-ylamine that tetracyclic tacrine analogues (Saracoglu, Kandemirli, 2008).

Recently, Chen *et al.* performed 3D-QSAR analysis on multi-target-directed AChEIs of tacrine–nimodipine dihydropyridine derivatives by using both CoMFA and CoMSIA methods (Chen, Liu, Zhao, & Zhang, 2012). The developed models were statistically significant and highly predictable (receptor based research CoMFA model, $q^2 = 0.686$, $r^2 = 0.948$; CoMSIA model, $q^2 = 0.756$, $r^2 = 0.907$ in all of the models no. of components = 6. Comparison of the methods for the predictabilities of QSAR models based on both aligns database and autodocking alignment for same test set (12 compounds).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Yan *et al.* developed the QSAR models using MLR analysis and SVM methods for predicting the inhibitory activity of 404 AChEIs (Yan, & Wang, 2012). The statistical analysis of Multi linear Regression Models and Support Vector Machine Models were analysed and suggested key chemical properties: molecule charge, electronegativity and H-bond donors for modeling and bioactivity of AChEIs. Recently, Jiang *et al.* has carried out CoMFA, CoMSIA, and hologram quantitative structure-activity relationship (HQSAR) analysis on AChEIs (Jiang, Yang, Chen, & Liang, 2013). Three different alignment rules were used to generate CoMFA and CoMSIA models to explore the effect of alignment methods on models. The developed models were statistically significant and highly predictable (CoMFA $q^2 = 0.748$, $r^2 = 0.996$, predicted $r^2 = 0.789$; CoMSIA $q^2 = 0.755$, $r^2 = 0.973$, predicted $r^2 = 0.706$; HQSAR $q^2 = 0.884$, $r^2 = 0.973$, predicted $r^2 = 0.734$) and revealed a considerable predictive ability.

Prado-Prado *et al.* used 3D MARCH-INSIDE (MI)-DRAGON, software to calculate 3D structural parameters (Prado-Prado García-Mera, Escobar, Alonso, Caamaño, Yañez, & González-Díaz, 2012). Both the classes of 3D parameters were used as input to train Artificial Neural Network (ANN) algorithms. This method was successfully used for the prediction of 22 different rasagiline derivatives with possible AChE inhibitory activity. Korabecny *et al.* have presented the synthesis and QSAR study of some novel series of 7-methoxytacrine (7-MEOTA)-donepezil like compounds for their ability to inhibit AChE and BuChE. The 7-MEOTA unit has less toxicity than tacrine (THA) derivative, connected with the analogues of N-benzylpiperazine moieties mimicking N-benzylpiperidine fragment from donepezil. 7-MEOTA-donepezil like compounds exerted mostly non-selective profile in inhibiting cholinesterases of different origin and is capable to simultaneously bind CS as well as PAS of AChE (Korabecny, Dolezal, Cabelova, Horova, Hruba, Ricny, Sedlacek, Nepovimova, Spilovska, Andrs, Musilek, Opletalova, Sepsova, Ripova, & Kuca, 2014).

Recently, Wong *et al.* reviewed the application of multi-criteria QSAR on AChEIs. It includes QSAR models for different series of compounds, comparative studies, and advances in methodologies (Wong, Duchowicz, Mercader, & Castro, 2012). SVM method was used for predicting AChEIs potential of unknown sets of compounds and for exhibiting the molecular descriptors associated with AChEIs (Lv and Xue, 2010).

Our research group has developed molecular docking guided comparative GFA, G/PLS, SVM and ANN QSAR models of structurally diverse 42 dual binding site AChEIs (Gupta *et al.* 2011). The binding conformations of 42 AChEIs were identified on the basis of the molecular docking approach as presented in Figure 18. These 42 dual binding site AChEIs were used in developing four different chemometric models using thermodynamic, electrotopological and electronic descriptors, in order to understand its structure-activity relationships, and thereby utilize in rational drug design approach. These compounds were partitioned into a training set of 31 compounds, and a test set of 11 compounds (75%: 25% ratio), in accordance with the HCA-Ward's algorithm.

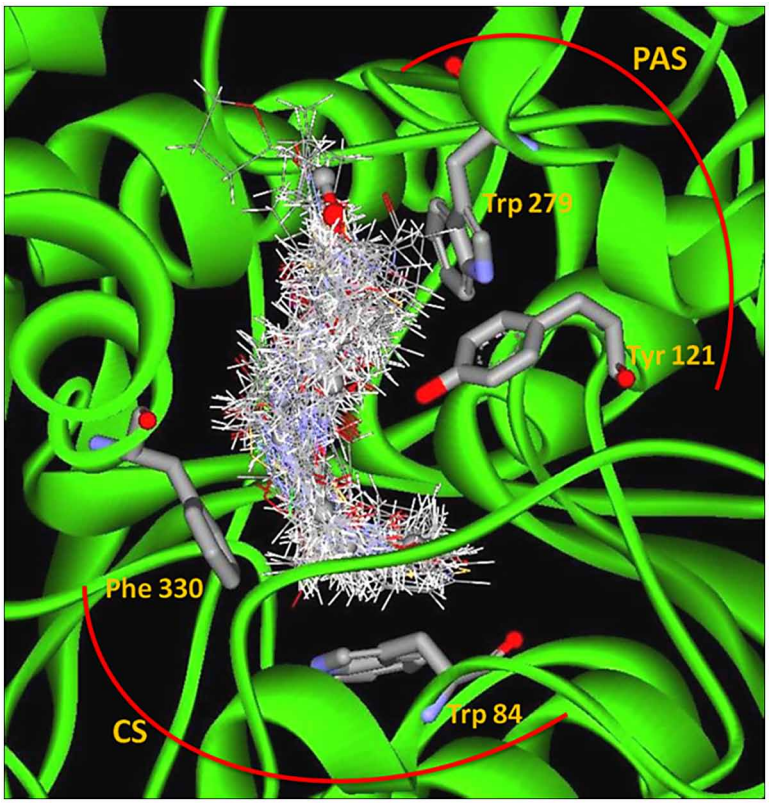
The GFA and G/PLS methods were used for variable selection as presented in model **6-7**. Best models were selected by the statistical measures, such as, squared correlation coefficient (R^2), cross-validated correlation coefficient (q^2), and predictive correlation coefficient R^2_{pred}, R^2 is a relative measure of the fit by the regression equation.

GFA Model

$$pIC_{50} = 5.793 + 0.443Atype_C_3 + 0.301Atype_C_32 - 0.965Atype_N_70 - 1.057S_ssS + 0.071Atype_H_46 - 0.215LUMO \quad (6)$$

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 18. Docked conformation of 42 dual binding site (CS and PAS) AChEIs (Gupta et al. 2011, 2012).



$N = 31$, $R^2 = 0.887$, LOF = 0.079, $R_a^2 = 0.859$, $q^2 = 0.838$, F = 31.44,

$n_{Test} = 11$, $R_{pred}^2 = 0.756$, $r^2 = 0.788$, $R_0^2 = 0.705$, $R_m^2 = 0.561$.

G/PLS Model

$$pIC_{50} = 5.808 + 0.469Atype_C_3 + 0.200Atype_N_75 - 1.446Atype_C_30 - 1.059S_ssS + 0.077Atype_H_46 - 0.209LUMO \quad (7)$$

$N = 31$, $R^2 = 0.889$, LSE = 0.029, $R_a^2 = 0.827$, $q^2 = 0.739$, F = 40.04,

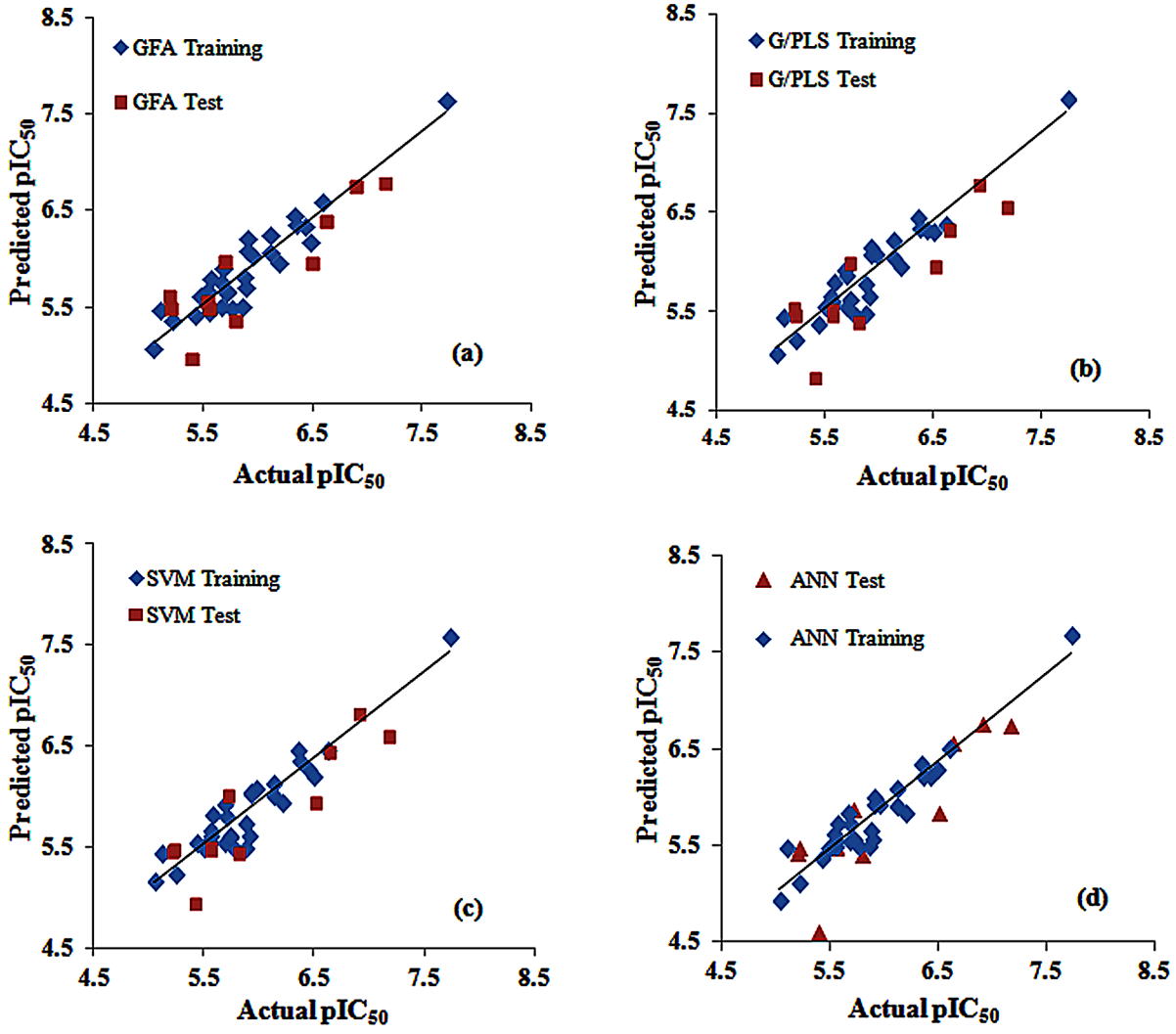
$n_{Test} = 11$, $R_{pred}^2 = 0.706$, $r^2 = 0.757$, $R_0^2 = 0.659$, $R_m^2 = 0.520$.

The values between the predicted versus experimental pIC_{50} of the training data set and the test data set compounds based on GFA, G/PLS, SVM and ANN QSAR models are presented in the scatter plot

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

as shown in Figure 19. The positive coefficient of atom-type thermodynamic descriptors, Atype_C_3, Atype_C_32 and Atype_H_46 in QSAR model obtained by GFA, indicated that this arrangement of carbon and hydrogen atoms in the compound increases the AChE inhibitory activity. The negative slope Atype_N_70 descriptor in GFA derived QSAR model revealed that AChE inhibitory activity decreases with an increase in the hydrophobicity associated with this nitrogen. The positive slope of the Atype_N_75 descriptor in QSAR model obtained from G/PLS revealed that AChE inhibitory activity increases with an increase in hydrophobicity associated with this nitrogen. The negative correlation of the Atype_C_30 descriptor in building QSAR model from G/PLS revealed that, for most potent compounds, its contribution

Figure 19. Scatter plot of actual vs. predicted activity of the training set (diamonds) and test set (triangles) compounds of QSAR model obtained using GFA method; (a), QSAR model obtained using the G/PLS method (b), QSAR model obtained using the SVM method (c), and QSAR model obtained using the ANN method (d)
(Gupta et al. 2011, 2012).



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

(or descriptor value) should be less, and for the least potent compounds its contribution should be more. Both GFA and G/PLS derived QSAR models showed negative correlation to S_ssS and LUMO descriptors. The negative coefficient of S_ssS (E-state index of the fragment of sulphur atom bonded with two single bonds) indicates that compounds having high S_ssS value have lower AChE inhibitory activity.

LUMO (lowest unoccupied molecular orbital energy), is an electronic descriptor, and the LUMO energy is the inverse measure of the electron accepting ability of a compound. In other words, higher the LUMO energy the lower will be the electron accepting ability of the compound. The above mentioned thermodynamic descriptors indicates that hydrophobicity of the atoms in different structural environment was essential for the better AChE inhibitory activity and which in turn enhances the brain uptake of drugs.

The robustness and significance of QSAR models were critically assessed using different cross-validation techniques on the test data set as presented in Table 2. The validation data were not involved by any means in the process of selecting the most appropriate descriptors or in the development of the QSAR models. Predictive ability (R^2_{pred}) of the selected models was evaluated by a test set of 11 compounds, and which are not included in the training data set.

The generated QSAR models using thermodynamic, electrotopological and electronic descriptors showed that non-linear methods are more robust than linear methods, and provide insight into the structural features of the compounds that are important for AChE inhibition.

Bolognesi *et al.* has done pioneering work on synthesis and molecular modeling for designing multipotent anti-Alzheimer drugs (Bolognesi, Cavalli, Valgimigli, Bartolini, Rosini, Andrisano, Recanatini, & Melchiorre, 2007). Molecular modeling, synthesis and kinetic evaluation of 11H-Indeno-[1,2-b]-quinolin-10-ylamine derivatives were reported by Rampa *et al.* (Rampa, Bisi, Belluti, Gobbi, Valenti, Andrisano, Cavrini, Cavalli, & Recanatini, 2000). The virtual screening study has been performed for lead identification of AChEIs–Histamine H_3 receptor antagonists from pharmacophore mapping, and further verified by QM/MM and docking methodology by Bembenek *et al.* (2008). Taft *et al.* have designed novel hybrid AChEIs by using molecular modeling, docking and ADMET studies (da Silva, Campo, Carvalho, & Taft, 2006). Mizutani *et al.* performed high-throughput virtual screening through docking analysis, for the discovery of novel AChEIs (Mizutani, & Itai, 2004). Recently, Ambure *et al.* has performed pharmacophore mapping-based virtual screening followed by molecular docking studies in search of potential AChEIs for the treatment of AD (Ambure, Kar & Roy, 2014).

Table 2. Leave one out cross validation parameter of QSAR models obtained from GFA and G/PLS, leave five out cross validation parameter of QSAR models obtained from SVM and ANN methodology

QSAR Models	R^2 (Training)	R^2_{LOO}	PRESS	S_{PRESS}	RMSE
GFA	0.887	[a]0.868	1.068	0.211	0.186
G/PLS	0.889	[a]0.849	1.221	0.256	0.198
SVM	0.885	[b]0.787	1.815	0.275	0.242
ANN	0.906	[b]0.795	1.697	0.266	0.234

[a]Leave one out method (LOO),[b]Leave five out method (L5O).
(Gupta *et al.* 2011, 2012).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

CONCLUSION AND FUTURE PROSPECTIVE

There has been a huge effort invested towards the discovery of new AChEIs during the past few decades. Several chemical classes of AChEIs were discovered and among them few have undergone for clinical trials during this period. The search of human AChEIs continues with the aim of identifying drugs with a better biological/chemical and low toxicity profile to currently marketed drugs. The examination of the quantitative models for the structure-activity relationships of several series of compounds belonging to diverse classes of AChEIs allowed us to understand the physico-chemical properties involved in the QSAR analysis. The influences of molecular descriptors are very promising in identifying the strong and weak inhibitors of diverse structural classes required for AChE enzyme inhibition. The development of QSAR models and computational prediction of biological activity will continue to be one of the most exciting areas of QSAR studies.

Here, we presented an overview of different QSAR methods and recent developments in AChEIs by different research groups. This is an attempt of consolidation of known data, with detailed knowledge about AChE inhibitor-enzyme binding modes, and structural insights gained from QSAR models, which might offer a starting point for the design of new more potent and safer AChEIs. The QSAR methodology can greatly help in this effort, being firmly founded on physico-chemical and statistical bases. In rational drug design, these are the requirements that must be met even more strictly in fields like that of the agent for the treatment of AD.

REFERENCES

Akamatsu, M. (2002). Current state and perspectives of 3D-QSAR. *Current Topics in Medicinal Chemistry*, *2*(12), 1381–1394. doi:10.2174/1568026023392887 PMID:12470286

Akula, N., Lecanu, L., Greeson, J., & Papadopoulos, V. (2006). 3D QSAR studies of AChE inhibitors based on molecular docking scores and CoMFA. *Bioorganic & Medicinal Chemistry Letters*, *16*(24), 6277–6280. doi:10.1016/j.bmcl.2006.09.030 PMID:17049234

Ambure, P., Kar, S., & Roy, K. (2014). Pharmacophore mapping-based virtual screening followed by molecular docking studies in search of potential acetylcholinesterase inhibitors as anti-Alzheimer's agents. *Bio Systems*, *116*, 10–20. doi:10.1016/j.biosystems.2013.12.002 PMID:24325852

Anderson, A. C. (2003). The process of structure-based drug design. *Chemistry & Biology*, *10*(9), 787–797. doi:10.1016/j.chembiol.2003.09.002 PMID:14522049

Asadabadi, E. B., Abdolmaleki, P., Barkooie, S. M. H., Jahandideh, S., & Rezaei, M. A. (2009). A combinatorial feature selection approach to describe the QSAR of dual site inhibitors of acetylcholinesterase. *Computers in Biology and Medicine*, *39*(12), 1089–1095. doi:10.1016/j.compbiomed.2009.09.003 PMID:19854437

Barnard, E. A. (1974). The peripheral nervous system. New York: Plenum Press.

Bartolini, M., Bertucci, C., Cavrini, V., & Andrisano, V. (2003). [beta]-Amyloid aggregation induced by human acetylcholinesterase: Inhibition studies. *Biochemical Pharmacology*, *65*(3), 407–416. doi:10.1016/S0006-2952(02)01514-9 PMID:12527333

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Bartus, R. T., Dean, R. L., Beer, B., & Lippa, A. S. (1982). The cholinergic hypothesis of geriatric memory dysfunction. *Science*, *217*(4558), 408–414. doi:10.1126/science.7046051 PMID:7046051

Baumann, K. (2003). Cross-validation as the objective function for variable-selection techniques. *Trends in Analytical Chemistry*, *22*(6), 395–406. doi:10.1016/S0165-9936(03)00607-1

Bembenek, S. D., Keith, J. M., Letavic, M. A., Apodaca, R., Barbier, A. J., & Dvorak, L. et al. (2008). Lead identification of acetylcholinesterase inhibitors–histamine H3 receptor antagonists from molecular modeling. *Bioorganic & Medicinal Chemistry*, *16*(6), 2968–2973. doi:10.1016/j.bmc.2007.12.048 PMID:18249544

Bencsura, A., Enyedy, I. Y., & Kovach, I. M. (1996). Probing the active site of acetylcholinesterase by molecular dynamics of its phosphonate ester adducts. *Journal of the American Chemical Society*, *118*(36), 8531–8541. doi:10.1021/ja952406v

Bermúdez-Lugo, J. A., Rosales-Hernández, M. C., Deeb, O., Trujillo-Ferrara, J., & Correa-Basurto, J. (2011). In silico methods to assist drug developers in acetylcholinesterase inhibitors design. *Current Medicinal Chemistry*, *18*(8), 1122–1136. doi:10.2174/092986711795029681 PMID:21291371

Bernard, P., Kireev, D. B., Chrétien, J. R., Fortier, P. L., & Coppet, L. (1999). Automated docking of 82 N-benzylpiperidine derivatives to mouse acetylcholinesterase and comparative molecular field analysis with'natural'alignment. *Journal of Computer-Aided Molecular Design*, *13*(4), 355–371. doi:10.1023/A:1008071118697 PMID:10425601

Bolognesi, M. L., Cavalli, A., Valgimigli, L., Bartolini, M., Rosini, M., & Andrisano, V. et al. (2007). Multi-target-directed drug design strategy: From a dual binding site acetylcholinesterase inhibitor to a trifunctional compound against Alzheimer's disease. *Journal of Medicinal Chemistry*, *50*(26), 6446–6449. doi:10.1021/jm701225u PMID:18047264

Butters, N., Delis, D. C., & Lucas, J. A. (1995). Clinical assessment of memory disorders in amnesia and dementia. *Annual Review of Psychology*, *46*(1), 493–523. doi:10.1146/annurev.ps.46.020195.002425 PMID:7872736

Camps, P., & Munoz-Torrero, D. (2002). Cholinergic drugs in pharmacotherapy of Alzheimer's disease. *Mini Reviews in Medicinal Chemistry*, *2*(1), 11–25. doi:10.2174/1389557023406638 PMID:12369954

Chaudhaery, S. S., Roy, K. K., & Saxena, A. K. (2009). Consensus superiority of the pharmacophore-based alignment, over maximum common substructure (MCS): 3D-QSAR studies on carbamates as acetylcholinesterase inhibitors. *Journal of Chemical Information and Modeling*, *49*(6), 1590–1601. doi:10.1021/ci900049e PMID:19441865

Chen, N., Liu, C., Zhao, L., & Zhang, H. (2012). 3D-QSAR study of multitarget- directed AChE inhibitors based on autodocking. *Medicinal Chemistry Research*, *21*(2), 245–256. doi:10.1007/s00044-010-9516-x

Cho, S. J., Garsia, M. L. S., Bier, J., & Tropsha, A. (1996). A. structure-based alignment and comparative molecular field analysis of acetylcholinesterase inhibitors. *Journal of Medicinal Chemistry*, *39*(26), 5064–5071. doi:10.1021/jm950771r PMID:8978837

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Choudhary, M. I., & Rode, B. M. (2003). 3D-QSAR studies on natural acetylcholinesterase inhibitors of sarcococca saligna by comparative molecular field analysis (CoMFA). *Bioorganic & Medicinal Chemistry Letters*, *13*(24), 4375–4380. doi:10.1016/j.bmcl.2003.09.034 PMID:14643329

Colombres, M., Sagal, J. P., & Inestrosa, N. C. (2004). An overview of the current and novel drugs for Alzheimer's disease with particular reference to anti-cholinesterase compounds. *Current Pharmaceutical Design*, *10*(25), 3121–3130. doi:10.2174/1381612043383359 PMID:15544502

Correa-Basurto, J., Flores-Sandoval, C., Marin-Cruz, J., Rojo-Dominguez, A., Espinoza-Fonseca, L. M., & Trujillo-Ferrara, J. G. (2007). Docking and quantum mechanics studies on cholinesterases and their inhibitors. *European Journal of Medicinal Chemistry*, *42*(1), 10–19. doi:10.1016/j.ejmech.2006.08.015 PMID:17055616

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. doi:10.1007/BF00994018

Cramer Iii, R. D., Patterson, D. E., & Bunce, J. D. (1988). Effect of shape on binding of steroids to carrier proteins. *Journal of the American Chemical Society*, *110*(18), 5959–5967. doi:10.1021/ja00226a005 PMID:22148765

Crivori, P., Cruciani, G., Carrupt, P. A., & Testa, B. (2000). Predicting blood-brain barrier permeation from three-dimensional molecular structure. *Journal of Medicinal Chemistry*, *43*(11), 2204–2216. doi:10.1021/jm990968+ PMID:10841799

Crum-Brown, A., & Fraser, T. R. (1868). On the connection between chemical constitution and physiological action. Part 1. On the physiological action of the salts of the ammonium bases, derived from Strychnia, Brucia, Thebia, Codeia, Morphia, and Nicotia. *Transactions of the Royal Society of Edinburgh*, *25*(01), 151–203. doi:10.1017/S0080456800028155

da Silva, C., Campo, V. L., Carvalho, I., & Taft, C. A. (2006). Molecular modeling, docking and ADMET studies applied to the design of a novel hybrid for treatment of Alzheimer's disease. *Journal of Molecular Graphics & Modelling*, *25*(2), 169–175. doi:10.1016/j.jmgm.2005.12.002 PMID:16413803

De Ferrari, G. V., Canales, M. A., Shin, I., Weiner, L. M., Silman, I., & Inestrosa, N. C. (2001). A structural motif of acetylcholinesterase that promotes amyloid-peptide fibril formation. *Biochemistry*, *40*(35), 447–410. doi:10.1021/bi0101392 PMID:11523986

Decker, M., Kraus, B., & Heilmann, J. (2008). Design, synthesis and pharmacological evaluation of hybrid molecules out of quinazolinimines and lipoic acid lead to highly potent and selective butyrylcholinesterase inhibitors with antioxidant properties. *Bioorganic & Medicinal Chemistry*, *16*(8), 4252–4261. doi:10.1016/j.bmc.2008.02.083 PMID:18343673

Dudek, A. Z., Arodz, T., & Galvez, J. (2006). Computational methods in developing quantitative structure-activity relationships (QSAR): A review. *Combinatorial Chemistry & High Throughput Screening*, *9*(3), 213–228. doi:10.2174/138620706776055539 PMID:16533155

Dunn Iii, W. J., Scott, D. R., & Glen, W. G. (1989). Principal components analysis and partial least squares regression. *Tetrahedron Comput. Methodol.*, *2*(6), 349–376. doi:10.1016/0898-5529(89)90004-3

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Fallarero, A., Oinonen, P., Gupta, S., Blom, P., Galkin, A., Mohan, C. G., & Vuorela, P. M. (2008). Inhibition of acetylcholinesterase by coumarins: The case of coumarin 106. *Pharmacological Research*, *58*(3-4), 215–221. doi:10.1016/j.phrs.2008.08.001 PMID:18778776

Friedman, J. H. (1991). Multivariate adaptive regression splines. *Annals of Statistics*, *19*(1), 1–141. doi:10.1214/aos/1176347963

Froede, H. C., & Wilson, I. B. (1971). Acetylcholinesterase. In P. D. Boyer (Ed.), *The enzymes* (3rd ed.; Vol. 5, pp. 87–114). New York: Academic Press.

Fuchs, S., Gurari, D., & Silman, I. (1974). Chemical modification of electric eel acetylcholinesterase by tetranitromethane. *Archives of Biochemistry and Biophysics*, *165*(1), 90–97. doi:10.1016/0003-9861(74)90145-3 PMID:4441088

Fuxreiter, M., & Warshel, A. (1998). Origin of the catalytic power of acetylcholinesterase: Computer simulation studies. *Journal of the American Chemical Society*, *120*(1), 183–194. doi:10.1021/ja972326m

Garg, D., Gandhi, T., & Mohan, C. G. (2008). Exploring QSTR and toxicophore of hERG K+ channel blockers using GFA and HypoGen techniques. *Journal of Molecular Graphics & Modelling*, *26*(6), 966–976. doi:10.1016/j.jmgm.2007.08.002 PMID:17928249

Gasteiger, J., & Zupan, J. (1993). Neural networks in chemistry. *Angewandte Chemie*, *32*(4), 503–527. doi:10.1002/anie.199305031

Geldenhuys, W. J., Lockman, P. R., Nguyen, T. H., Van der Schyf, C. J., Crooks, P. A., Dwoskin, L. P., & Allen, D. D. (2005). 3D-QSAR study of bis-azaaromatic quaternary ammonium analogs at the blood–brain barrier choline transporter. *Bioorganic & Medicinal Chemistry*, *13*(13), 4253–4261. doi:10.1016/j.bmc.2005.04.020 PMID:15878282

Goeldner, M. P., & Hirth, C. G. (1980). Specific photoaffinity labeling induced by energy transfer: Application to irreversible inhibition of acetylcholinesterase. *Proceedings of the National Academy of Sciences of the United States of America*, *77*(11), 6439–6442. doi:10.1073/pnas.77.11.6439 PMID:6935657

Golbraikh, A., & Tropsha, A. (2002). Beware of q2! *Journal of Molecular Graphics & Modelling*, *20*(4), 269–276. doi:10.1016/S1093-3263(01)00123-1 PMID:11858635

Gomeni, R., Bani, M., D'Angeli, C., Corsi, M., & Bye, A. (2001). Computer-assisted drug development (CADD): An emerging technology for designing first-time-in-man and proof-of-concept studies from preclinical experiments. *European Journal of Pharmaceutical Sciences*, *13*(3), 261–270. doi:10.1016/S0928-0987(01)00111-7 PMID:11384848

Greig, N. H., Utsuki, T., Ingram, D. K., Wang, Y., Pepeu, G., & Scali, C. et al. (2005). Selective butyrylcholinesterase inhibition elevates brain acetylcholine, augments learning and lowers Alzheimer beta-amyloid peptide in rodent. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(47), 17213–17218. doi:10.1073/pnas.0508575102 PMID:16275899

Gupta, S. (2012). *Computational study of dual binding site acetylcholinesterase inhibitors for Alzheimer's disease*. (Ph.D. Thesis). National Institute of Pharmaceutical Education and Research (NIPER), SAS Nagar, Punjab, India.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Gupta, S., Fallarero, A., Vainio, M. J., Saravanan, P., Santeri Puranen, J., & Järvinen, P. et al. (2011). Molecular docking guided comparative GFA, G/PLS, SVM and ANN models of structurally diverse dual binding site acetylcholinesterase inhibitors. *Mol. Inform.*, *30*(8), 689–706.

Gupta, S. P. (2007). Quantitative structure-activity relationship studies on zinc-containing metalloproteinase inhibitors. *Chemical Reviews*, *107*(7), 3042–3087. doi:10.1021/cr030448t PMID:17622180

Hammett, L. P. (1935). Some relations between reaction rates and equilibrium constants. *Chemical Reviews*, *17*(1), 125–136. doi:10.1021/cr60056a010

Hansch, C., & Fujita, T. (1964). p-σ-π Analysis. A method for the correlation of biological activity and chemical structure. *Journal of the American Chemical Society*, *86*(8), 1616–1626. doi:10.1021/ja01062a035

Hansch, C., Hoekman, D., & Gao, H. (1996). Comparative QSAR: Toward a deeper understanding of chemicobiological interactions. *Chemical Reviews*, *96*(3), 1045–1076. doi:10.1021/cr9400976 PMID:11848780

Hansch, C., Hoekman, D., Leo, A., Weininger, D., & Selassie, C. D. (2002). Chem-bioinformatics: Comparative QSAR at the interface between chemistry and biology. *Chemical Reviews*, *102*(3), 783–812. doi:10.1021/cr0102009 PMID:11890757

Hansch, C., Maloney, P. P., Fujita, T., & Muir, R. M. (1962). Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients. *Nature*, *194*(4824), 178–180. doi:10.1038/194178b0

Harel, M., Schalk, I., Ehret-Sabatier, L., Bouet, F., Goeldner, M., & Hirth, C. et al. (1993). Quaternary ligand binding to aromatic residues in the active-site gorge of acetylcholinesterase. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(19), 9031–9035. doi:10.1073/pnas.90.19.9031 PMID:8415649

Hasegawa, K., Kimura, T., & Funatsu, K. (1999). GA strategy for variable selection in QSAR studies: Application of GA-based region selection to a 3D-QSAR study of acetylcholinesterase inhibitors. *Journal of Chemical Information and Computer Sciences*, *39*(1), 112–120. doi:10.1021/ci980088o PMID:10094610

Heydorn, W. E. (1997). Donepezil (E2020): A new acetylcholinesterase inhibitor. Review of its pharmacology, pharmacokinetics, and utility in the treatment of Alzheimer's disease. *Expert Opinion on Investigational Drugs*, *6*(10), 1527–1535. doi:10.1517/13543784.6.10.1527 PMID:15989517

Hillisch, A., Pineda, L. F., & Hilgenfeld, R. (2004). Utility of homology models in the drug discovery process. *Drug Discovery Today*, *9*(15), 659–669. doi:10.1016/S1359-6446(04)03196-4 PMID:15279849

Holzgrabe, U., Kapkova, P., Alptüzün, V., Scheiber, J., & Kugelmann, E. (2007). Targeting acetylcholinesterase to treat neurodegeneration. *Expert Opinion on Therapeutic Targets*, *11*(2), 161–179. doi:10.1517/14728222.11.2.161 PMID:17227232

Hostettmann, K., Borloz, A., Urbain, A., & Marston, A. (2006). Natural product inhibitors of acetylcholinesterase. *Current Organic Chemistry*, *10*(8), 825–847. doi:10.2174/138527206776894410

Houghton, P. J., Ren, Y., & Howes, M. J. (2006). Acetylcholinesterase inhibitors from plants and fungi. *Natural Product Reports*, *23*(2), 181–199. doi:10.1039/b508966m PMID:16572227

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Hurley, M. M., Wright, J. B., Lushington, G. H., & White, W. E. (2003). Quantum mechanics and mixed quantum mechanics/molecular mechanics simulations of model nerve agents with acetylcholinesterase. *Theoretica Chimica Acta*, *109*, 160–168.

Inestrosa, N. C., Alvarez, A., Perez, C. A., Moreno, R. D., Vicente, M., & Linker, C. et al. (1996). Acetylcholinesterase accelerates assembly of amyloid-beta-peptides into Alzheimer's fibrils: Possible role of the peripheral site of the enzyme. *Neuron*, *16*(4), 881–891. doi:10.1016/S0896-6273(00)80108-7 PMID:8608006

Inestrosa, N. C., Dinamarca, M. C., Alvarez, A., & Center, C. B. (2008). Amyloid–cholinesterase interactions. *The FEBS Journal*, *275*(4), 625–632. doi:10.1111/j.1742-4658.2007.06238.x PMID:18205831

Jiang, Y. R., Yang, Y. Y., Chen, Y. L., & Liang, Z. J. (2013). CoMFA, CoMSIA and HQSAR studies of acetylcholinesterase inhibitors. *Curr. Comput. Aided Drug Des.*, *9*(3), 385–395. doi:10.2174/15734099113099990015 PMID:24010934

Jin, Y. P., Gatz, M., Johansson, B., & Pedersen, N. (2004). Sensitivity and specificity of dementia coding in two Swedish disease registries. *Neurology*, *63*(4), 739–741. doi:10.1212/01.WNL.0000134604.48018.97 PMID:15326258

John, G. C., Andrew, M. D., Sorel, M., Markus, H., & Hongming, C. (2013). Chemical predictive modelling to improve compound quality. *Nature Reviews. Drug Discovery*, *12*(12), 948–962. doi:10.1038/nrd4128 PMID:24287782

Jung, M., Tak, J., Lee, Y., & Jung, Y. (2007). Quantitative structure–activity relationship (QSAR) of tacrine derivatives against acetylcholinesterase (AChE) activity using variable selections. *Bioorganic & Medicinal Chemistry Letters*, *17*(4), 1082–1090. doi:10.1016/j.bmcl.2006.11.022 PMID:17158047

Kapetanovic, I. M. (2008). Computer-aided drug discovery and development (CADDD): In silico-chemico-biological approach. *Chemico-Biological Interactions*, *171*(2), 165–176. doi:10.1016/j.cbi.2006.12.006 PMID:17229415

Katritzky, A. R., Petrukhin, R., Yang, H., & Karelson, M. (2005). *CODESSA PRO*. Florida, USA.

Klebe, G., Abraham, U., & Mietzner, T. (1994). Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *Journal of Medicinal Chemistry*, *37*(24), 4130–4146. doi:10.1021/jm00050a010 PMID:7990113

Knapp, M. J., Knopman, D. S., Solomon, P. R., Pendlebury, W. W., Davis, C. S., & Gracon, S. I. (1994). A 30-week randomized controlled trial of high-dose tacrine in patients with Alzheimer's disease. The Tacrine study group. *Journal of the American Medical Association*, *271*(13), 985–991. doi:10.1001/jama.1994.03510370037029 PMID:8139083

Korabecny, J., Dolezal, R., Cabelova, P., Horova, A., Hruba, E., & Ricny, J. et al. (2014). 7-MEOTA-donepezil like compounds as cholinesterase inhibitors: Synthesis, pharmacological evaluation, molecular modeling and QSAR studies. *European Journal of Medicinal Chemistry*, *23*, 426–438. doi:10.1016/j.ejmech.2014.05.066 PMID:24929293

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Kryger, G., Harel, M., Giles, K., Toker, L., Velan, B., & Lazar, A. et al. (2000). Structures of recombinant native and E202Q mutant human acetylcholinesterase complexed with the snake-venom toxin fasciculin-II. *Acta Crystallogr. Sec. D*, *56*(11), 1385–1394. doi:10.1107/S0907444900010659 PMID:11053835

Kryger, G., Silman, I., & Sussman, J. L. (1999). Structure of acetylcholinesterase complexed with E2020 (Aricept®): Implications for the design of new anti-Alzheimer drugs. *Structure (London, England)*, *7*(3), 297–307. doi:10.1016/S0969-2126(99)80040-9 PMID:10368299

Kubinyi, H. (1997). QSAR and 3D QSAR in drug design Part 1: Methodology. *Drug Discovery Today*, *2*(11), 457–467. doi:10.1016/S1359-6446(97)01079-9

La Du, B. N. (1994). Human cholinesterases and anticholinesterases. *American Journal of Human Genetics*, *55*, 593–594.

Levitt, M., & Chothia, C. (1976). Structural patterns in globular proteins. *Nature*, *261*(5561), 552–558. doi:10.1038/261552a0 PMID:934293

Li, H., Yap, C. W., Ung, C. Y., Xue, Y., Li, Z. R., & Han, L. Y. et al. (2007). Machine learning approaches for predicting compounds that interact with therapeutic and ADMET related proteins. *Journal of Pharmaceutical Sciences*, *96*(11), 2838–2860. doi:10.1002/jps.20985 PMID:17786989

Lill, M. A. (2007). Multi-dimensional QSAR in drug discovery. *Drug Discovery Today*, *12*(23-24), 1013–1017. doi:10.1016/j.drudis.2007.08.004 PMID:18061879

Lv, W., & Xue, Y. (2010). Prediction of acetylcholinesterase inhibitors and characterization of correlative molecular descriptors by machine learning methods. *European Journal of Medicinal Chemistry*, *45*(3), 1167–1172. doi:10.1016/j.ejmech.2009.12.038 PMID:20053484

Malik, N. N. (2008). Drug discovery: Past, present and future. *Drug Discovery Today*, *13*(21-22), 909–912. doi:10.1016/j.drudis.2008.09.007 PMID:18852066

Martin-Santamaria, S., Munoz-Muriedas, J., Luque, F. J., & Gago, F. (2004). Modulation of binding strength in several classes of active site inhibitors of acetylcholinesterase studied by comparative binding energy analysis. *Journal of Medicinal Chemistry*, *47*(18), 4471–4482. doi:10.1021/jm049877p PMID:15317459

Martinez, A., Fernandez, E., Castro, A., Conde, S., Rodriguez-Franco, I., Baños, J. E., & Badia, A. (2000). N-Benzylpiperidine derivatives of 1,2,4-thiadiazolidinone as new acetylcholinesterase inhibitors. *European Journal of Medicinal Chemistry*, *35*(10), 913–922. doi:10.1016/S0223-5234(00)01166-1 PMID:11121617

Maulet, Y., Camp, S., Gibney, G., Rachinsky, T. L., Ekström, T. J., & Taylor, P. (1990). Single gene encodes glycophospholipid-anchored and asymmetric acetylcholinesterase forms: Alternative coding exons contain inverted repeat sequences. *Neuron*, *4*(2), 289–301. doi:10.1016/0896-6273(90)90103-M PMID:2306366

Mayeux, R., & Sano, M. (1999). Treatment of Alzheimer's disease. *The New England Journal of Medicine*, *341*(22), 1670–1679. doi:10.1056/NEJM199911253412207 PMID:10572156

Mekenyan, O., & Bonchev, D. (1986). OASIS method for predicting biological activity of chemical compounds. *Acta Pharmaceutica Jugoslavica*, *36*, 225–237.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Meyer, H. (1899). Theorie der alkoholnarkose. *Arch. Exp. Pathol. Pharmakol.*, *42*(2-4), 109–118. doi:10.1007/BF01834479

Mirjana, B. C., Danijela, Z. K., Tamara, D. L., Aleksandra, M. B., & Vesana, M. V. (2013). Acetylcholinesterase inhibitors: Pharmacology and toxicology. *Current Neuropharmacology*, *11*(3), 315–335. doi:10.2174/1570159X11311030006 PMID:24179466

Mitra, I., Saha, A., & Roy, K. (2010). QSPR of antioxidant phenolic compounds using quantum chemical descriptors. *Molecular Simulation*, *37*(5), 394–413. doi:10.1080/08927022.2010.543980

Mizutani, M. Y., & Itai, A. (2004). Efficient method for high-throughput virtual screening based on flexible docking: discovery of novel acetylcholinesterase inhibitors. *Journal of Medicinal Chemistry*, *47*(20), 4818–4828. doi:10.1021/jm030605g PMID:15369385

Mooser, G., & Sigman, D. S. (1974). Ligand binding properties of acetylcholinesterase determined with fluorescent probes. *Biochemistry*, *13*(11), 2299–2307. doi:10.1021/bi00708a010 PMID:4857568

Mullan, M., Crawford, F., Axelman, K., Houlden, H., Lilius, L., Winblad, B., & Lannfelt, L. (1992). A pathogenic mutation for probable Alzheimer's disease in the APP gene at the N-terminus of beta- amyloid. *Nature Genetics*, *1*(5), 345–347. doi:10.1038/ng0892-345 PMID:1302033

Naslund, J., Haroutunian, V., Mohs, R., Davis, K. L., Davies, P., Greengard, P., & Buxbaum, J. D. (2000). Correlation between elevated levels of amyloid beta-peptide in the brain and cognitive decline. *Journal of the American Medical Association*, *283*(12), 1571–1577. doi:10.1001/jama.283.12.1571 PMID:10735393

Novak, M., & Rajagopal, S. (2002). Correlations of nitrenium ion selectivities with quantitative mutagenicity and carcinogenicity of the corresponding amines. *Chemical Research in Toxicology*, *15*(12), 1495–1503. doi:10.1021/tx025584s PMID:12482231

Ollis, D. L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., & Franken, S. M. et al. (1992). The α/β hydrolase fold. *Protein Engineering*, *5*(3), 197–211. doi:10.1093/protein/5.3.197 PMID:1409539

Ooms, F. (2000). Molecular modeling and computer aided drug design. Examples of their applications in medicinal chemistry. *Current Medicinal Chemistry*, *7*(2), 141–158. doi:10.2174/0929867003375317 PMID:10637360

Overton, E. (1895). Über die osmotischen eigenschaften der lebenden pflanzen und tierzelle. *Vierteljahrsshriften Naturforschungen Ges Zurich*, *40*, 159–201.

Pang, Y. P., & Kozikowski, A. P. (1994). Prediction of the binding site of 1-benzyl-4-[(5, 6-dimethoxy-1-indanon-2-yl) methyl] piperidine in acetylcholinesterase by docking studies with the SYSDOC program. *Journal of Computer-Aided Molecular Design*, *8*(6), 683–693. doi:10.1007/BF00124015 PMID:7738604

Pang, Y. P., & Kozikowski, A. P. (1994). Prediction of the binding sites of huperzine A in acetylcholinesterase by docking studies. *Journal of Computer-Aided Molecular Design*, *8*(6), 669–681. doi:10.1007/BF00124014 PMID:7738603

Pang, Y. P., Quiram, P., Jelacic, T., Hong, F., & Brimijoin, S. (1996). Highly potent, selective, and low cost bis-tetrahydroaminacrine inhibitors of acetylcholinesterase. *The Journal of Biological Chemistry*, *271*(39), 23646–23649. doi:10.1074/jbc.271.39.23646 PMID:8798583

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Prado-Prado, F., García-Mera, X., Escobar, M., Alonso, N., Caamaño, O., Yañez, M., & González-Díaz, H. (2012). 3D MI-DRAGON: New model for the reconstruction of US FDA drug- target network and theoretical-experimental studies of inhibitors of rasagiline derivatives for AChE. *Current Topics in Medicinal Chemistry*, *12*(16), 1843–1865. doi:10.2174/156802612803989228 PMID:23030618

Quinn, D. M. (1987). Acetylcholinesterase: Enzyme structure, reaction dynamics, and virtual transition states. *Chemical Reviews*, *87*(5), 955–979. doi:10.1021/cr00081a005

Rampa, A., Bisi, A., Belluti, F., Gobbi, S., Valenti, P., & Andrisano, V. et al. (2000). Acetylcholinesterase inhibitors for potential use in Alzheimer's disease: Molecular modeling, synthesis and kinetic evaluation of 11H-indeno-[1,2-b]-quinolin-10-ylamine derivatives. *Bioorganic & Medicinal Chemistry*, *8*(3), 497–506. doi:10.1016/S0968-0896(99)00306-5 PMID:10732965

Recanatini, M., Cavalli, A., Belluti, F., Piazzi, L., Rampa, A., & Bisi, A. et al. (2000). SAR of 9-amino-1,2,3,4-tetrahydroacridine-based acetylcholinesterase inhibitors: Synthesis, enzyme inhibitory activity, QSAR, and structure-based CoMFA of tacrineanalogues. *Journal of Medicinal Chemistry*, *43*(10), 2007–2018. doi:10.1021/jm990971t PMID:10821713

Recanatini, M., Cavalli, A., & Hansch, C. A. (1997). Comparative QSAR analysis of acetylcholinesterase inhibitors currently studied for the treatment of Alzheimer's disease. *Chemico-Biological Interactions*, *105*(3), 199–228. doi:10.1016/S0009-2797(97)00047-1 PMID:9291997

Rogers, D., & Hopfinger, A. J. (1994). Application of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships. *Journal of Chemical Information and Computer Sciences*, *34*(4), 854–866. doi:10.1021/ci00020a020

Ros, E., Aleu, J., Marsal, J., & Solsona, C. (2000). Effects of CI-1002 and CI-1017 on spontaneous synaptic activity and on the nicotinic acetylcholine receptor of Torpedo electric organ. *European Journal of Pharmacology*, *390*(1-2), 7–13. doi:10.1016/S0014-2999(99)00911-5 PMID:10708701

Rosenberg, R. N. (2000). The molecular and genetic basis of AD: the end of the beginning: the Wartenberg Lecture. *Neurology*, *54*(11), 2045–2054. doi:10.1212/WNL.54.11.2045 PMID:10851361

Rosenberry, T. L. (1975). Acetylcholinesterase. *Adv. Enzymol*, *43*, 103–218. PMID:891

Roy, K. (2007). On some aspects of validation of predictive quantitative structure-activity relationship models. *Expert Opin. Drug Discov.*, *2*(12), 1567–1577. doi:10.1517/17460441.2.12.1567 PMID:23488901

Roy, K. K., Dixit, A., & Saxena, A. K. (2008). An investigation of structurally diverse carbamates for acetylcholinesterase (AChE) inhibition using 3D-QSAR analysis. *Journal of Molecular Graphics & Modelling*, *27*(2), 197–208. doi:10.1016/j.jmgm.2008.04.006 PMID:18515163

Roy, P. P., & Roy, K. (2008). On some aspects of variable selection for partial least squares regression models. *QSAR & Combinatorial Science*, *27*(3), 302–313. doi:10.1002/qsar.200710043

Rücker, C., Meringer, M., & Kerber, A. (2004). QSPR using MOLGEN-QSPR: The example of haloalkane boiling points. *Journal of Chemical Information and Computer Sciences*, *44*(6), 2070–2076. doi:10.1021/ci049802u PMID:15554677

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Rücker, C., Rücker, G., & Meringer, M. (2007). y-Randomization and its variants in QSPR/QSAR. *Journal of Chemical Information and Modeling*, *47*(6), 2345–2357. doi:10.1021/ci700157b PMID:17880194

Saracoglu, M., & Kandemirli, F. (2008). The investigation of structure activity relationships of tacrine analogues: Electronic-topological method. *Open Med. Chem. J.*, *2*, 75–78. doi:10.2174/1874104500802010075 PMID:19662147

Schneider, L. S. (2000). The future of cholinergic replacement therapy in Alzheimer's disease. *Current Opinion in Investigational Drugs (London, England)*, *2*, 427–437.

Shen, L., Liu, G., & Tang, Y. (2007). Molecular docking and 3D-QSAR studies of 2-substituted 1-indanone derivatives as acetylcholinesterase inhibitors. *Acta Pharmacologica Sinica*, *28*(12), 2053–2063. doi:10.1111/j.1745-7254.2007.00664.x PMID:18031622

Shutske, G. M., Pierrat, F. A., Kapples, K. J., Cornfelt, M. L., Szewczak, M. R., & Huger, F. P. et al. (1989). 9-Amino-1,2,3,4-tetrahydroacridin-1-ols: Synthesis and evaluation as potential Alzheimer's disease therapeutics. *Journal of Medicinal Chemistry*, *32*(8), 1805–1813. doi:10.1021/jm00128a024 PMID:2754707

Silmana, I., & Sussman, J. L. (2008). Acetylcholinesterase: How is structure related to function? *Chemico-Biological Interactions*, *175*(1-3), 3–10. doi:10.1016/j.cbi.2008.05.035 PMID:18586019

Silverman, B. D., & Daniel, E. P. (1996). Comparative molecular moment analysis (CoMMA): 3D-QSAR without molecular superposition. *Journal of Medicinal Chemistry*, *39*(11), 2129–2140. doi:10.1021/jm950589q PMID:8667357

Sippl, W., Contreras, J. M., Parrot, I., Rival, Y. M., & Wermuth, C. G. (2001). Structure-based 3D QSAR and design of novel acetylcholinesterase inhibitors. *Journal of Computer-Aided Molecular Design*, *15*(5), 395–410. doi:10.1023/A:1011150215288 PMID:11394735

Song, C. M., Lim, S. J., & Tong, J. C. (2009). Recent advances in computer-aided drug design. *Briefings in Bioinformatics*, *10*(5), 579–591. doi:10.1093/bib/bbp023 PMID:19433475

Sramek, J. J., Frackiewicz, E. J., & Cutler, N. R. (2000). Review of the acetylcholinesterase inhibitor galanthamine. *Expert Opinion on Investigational Drugs*, *9*(10), 2393–2402. doi:10.1517/13543784.9.10.2393 PMID:11060814

Stahl, S. M. (1999). Molecular neurobiology for practicing psychiatrists, Part 4: Transferring the message of chemical neurotransmission from presynaptic neurotransmitter to postsynaptic gene expression. *The Journal of Clinical Psychiatry*, *60*(12), 813–814. doi:10.4088/JCP.v60n1201 PMID:10665625

Stuper, A. J., & Jurs, P. C. (1976). ADAPT: A computer system for automated data analysis using pattern recognition techniques. *Journal of Chemical Information and Computer Sciences*, *16*(2), 99–105. doi:10.1021/ci60006a014

Sugimoto, H., Yamanish, Y., Iimura, Y., & Kawakami, Y. (2000). Donepezil hydrochloride (E2020) and other acetylcholinesterase inhibitors. *Current Medicinal Chemistry*, *7*(3), 303–339. doi:10.2174/0929867003375191 PMID:10637367

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Sulea, T., Kurunczi, L., Oprea, T. I., & Simon, Z. (1998). MTD-ADJ: A multiconformational minimal topologic difference for determining bioactive conformers using adjusted biological activities. *Journal of Computer-Aided Molecular Design*, *12*(2), 133–146. doi:10.1023/A:1007913622673 PMID:9690173

Sussman, J. L., Harel, M., Frolow, F., Oefner, C., Goldman, A., Toker, L., & Silman, I. (1991). Atomic structure of acetylcholinesterase from Torpedo californica: A prototypic acetylcholine-binding protein. *Science*, *253*(5022), 872–879. doi:10.1126/science.1678899 PMID:1678899

Sutter, J. M., & Jurs, P. C. (1996). Prediction of aqueous solubility for a diverse set of heteroatom-containing organic compounds using a quantitative structure-property relationship. *Journal of Chemical Information and Computer Sciences*, *36*(1), 100–107. doi:10.1021/ci9501507

Taft, C. A., da Silva, V. B., & da Silva, C. H. T. (2008). Current topics in computer aided drug design. *Journal of Pharmaceutical Sciences*, *97*(3), 1089–1098. doi:10.1002/jps.21293 PMID:18214973

Taft, R. W. Jr. (1952). Polar and steric substituent constants for aliphatic and o-Benzoate groups from rates of esterification and hydrolysis of esters. *Journal of the American Chemical Society*, *74*(12), 3120–3128. doi:10.1021/ja01132a049

Tanzi, R. E., Kovacs, D. M., Kim, T. W., Moir, R. D., Guenette, S. Y., & Wasco, W. (1996). The gene defects responsible for familial Alzheimer's disease. *Neurobiology of Disease*, *3*(3), 159–168. doi:10.1006/nbdi.1996.0016 PMID:8980016

Taylor, P., & Radic, Z. (1994). The cholinesterases: From genes to proteins. *Annual Review of Pharmacology and Toxicology*, *34*(1), 281–320. doi:10.1146/annurev.pa.34.040194.001433 PMID:8042853

Todeschini, R., Consonni, V., Mauri, A., & Pavan, M. (2003). Software dragon: Calculation of molecular descriptors. Department of Environmental Sciences, University of Milano-Bicocca, and Talete, srl.

Todeschini, R., Consonni, V., Mauri, A., & Pavan, M. (2004). Detecting "bad" regression models: Multicriteria fitness functions in regression analysis. *Analytica Chimica Acta*, *515*(1), 199–208. doi:10.1016/j.aca.2003.12.010

Tong, W., Collantes, E. R., Chen, Y., & Welsh, W. J. (1996). A comparative molecular field analysis study of N-benzylpiperidines as acetylcholinesterase inhibitors. *Journal of Medicinal Chemistry*, *39*(2), 380–387. doi:10.1021/jm950704x PMID:8558505

Tropsha, A., Gramatica, P., & Gombar, V. K. (2003). The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR & Combinatorial Science*, *22*(1), 69–77. doi:10.1002/qsar.200390007

Vasilyev, V. V. (1994). Tetrahedral intermediate formation in the acylation step of acetylcholinesterases. A combined quantum chemical and molecular mechanical model. *Journal of Molecular Structure*, *304*(2), 129–141. doi:10.1016/S0166-1280(96)80005-4

Villalobos, A., Blake, J. F., Biggers, C. K., Butler, T. W., Chapin, D. S., & Chen, Y. L. et al. (1994). Novel benzisoxazole derivatives as potent and selective inhibitors of acetylcholinesterase. *Journal of Medicinal Chemistry*, *37*(17), 2721–2734. doi:10.1021/jm00043a012 PMID:8064800

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Wang, Q. M., Jiang, H. L., Chen, K. X., Ji, R. Y., & Ye, Y. J. (1999). Theoretical studies on the possible reaction pathway for the deacylation of the AChE catalyzed reaction. *International Journal of Quantum Chemistry*, *74*(3), 315–325. doi:10.1002/(SICI)1097-461X(1999)74:3<315::AID-QUA4>3.0.CO;2-Y

Wang, Y., Chiu, J. F., & He, Q. Y. (2005). Proteomics in computer-aided drug design. *Curr. Comput. Aided Drug Des.*, *1*(1), 43–52. doi:10.2174/1573409052952260

Weise, C., Kreienkamp, H. J., Raba, R., Pedak, A., Aaviksaar, A., & Hucho, F. (1990). Anionic subsites of the acetylcholinesterase from Torpedo californica: Affinity labelling with the cationic reagent N, N-dimethyl-2-phenyl-aziridinium. *The EMBO Journal*, *9*, 3885–3888. PMID:2249655

Winkler, D. A. (2002). The role of quantitative structure - activity relationships (QSAR) in biomolecular discovery. *Briefings in Bioinformatics*, *3*(1), 73–86. doi:10.1093/bib/3.1.73 PMID:12002226

Wold, S., & Eriksson, L. (1995). Chemometric methods. In H. van de Waterbeemd (Ed.), Molecular design (pp. 312–317). Weinheim, Germany: VCH.

Wong, K. Y., Duchowicz, P. R., Mercader, A. G., & Castro, E. A. (2012). QSAR applications during last decade on inhibitors of acetylcholinesterase in Alzheimer's disease. *Mini Reviews in Medicinal Chemistry*, *12*(10), 936–946. doi:10.2174/138955712802762365 PMID:22303974

Xiao, X. Q., Yang, J. W., & Tang, X. C. (1999). Huperzine A protects rat pheochromocytoma cells against hydrogen peroxide-induced injury. *Neuroscience Letters*, *275*(2), 73–76. doi:10.1016/S0304-3940(99)00695-3 PMID:10568502

Yan, A., & Wang, K. (2012). Quantitative structure and bioactivity relationship study on human acetylcholinesterase inhibitors. *Bioorganic & Medicinal Chemistry Letters*, *22*(9), 3336–3342. doi:10.1016/j.bmcl.2012.02.108 PMID:22460031

Yap, C. W., Li, H., Ji, Z. L., & Chen, Y. Z. (2007). Regression methods for developing QSAR and QSPR models to predict compounds of specific pharmacodynamic, pharmacokinetic and toxicological properties. *Mini Reviews in Medicinal Chemistry*, *7*(11), 1097–1107. doi:10.2174/138955707782331696 PMID:18045213

Yuan, Y., Zhang, R., & Luo, L. (2009). Classification study of novel piperazines as antagonists for the melanocortin-4 receptor based on least-squares support vector machines. *Chemometrics and Intelligent Laboratory Systems*, *96*(2), 144–148. doi:10.1016/j.chemolab.2009.01.004

Zhang, Y., Kua, J., & McCammon, J. A. (2002). Role of the catalytic triad and oxyanion hole in acetylcholinesterase catalysis: An ab initio QM/MM study. *Journal of the American Chemical Society*, *124*(35), 10572–10577. doi:10.1021/ja020243m PMID:12197759

Zhou, Y., Wang, S., & Zhang, Y. (2010). Catalytic reaction mechanism of acetylcholinesterase determined by Born- Oppenheimer *Ab-Initio* QM/MM molecular dynamics simulations. *The Journal of Physical Chemistry B*, *114*(26), 8817–8825. doi:10.1021/jp104258d PMID:20550161

Zupan, J. (1994). Introduction to artificial neural network (ANN) methods: What they are and how to use them. *Acta Chimica Slovenica*, *41*, 327–327.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

KEY TERMS AND DEFINITIONS

AChE Enzyme: An enzyme that catalyses the hydrolysis of acetylcholine to acetate and choline.

Alzheimer's Disease: It is a neurodegenerative disorder coined by German psychiatrist and neuropathologist Alois Alzheimer in 1906.

GFA: Genetic function approximation is genetics based linear regression method of variable selection that combines Holland's genetic algorithm with Friedman's multivariate adaptive regression splines.

In Silico: A Latin type expression and define the work performed using computer simulation or computer modeling.

QSAR: Quantitative structure–activity relationship models are statistical regression equations used in the chemical, biological and engineering sciences. Model influences the chemical structure, descriptor, physico-chemical properties on different physical, chemical and biological endpoints, as well as in understanding the chemical processes.

This work was previously published in Quantitative Structure-Activity Relationships in Drug Design, Predictive Toxicology, and Risk Assessment edited by Kunal Roy, pages 354-399, copyright year 2015 by Medical Information Science Reference (an imprint of IGI Global).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 23
Ethics, Risk, and Media Intervention:
Women's Breast Cancer in Venezuela

Mahmoud Eid
University of Ottawa, Canada

Isaac Nahon-Serfaty
University of Ottawa, Canada

ABSTRACT

Breast cancer incidence and mortality rates are of concern among Latin American women, mainly due to the growing prevalence of this disease and the lack of compliance to proper breast cancer screening and treatment. Focusing on Venezuelan women and the challenges and barriers that interact with their health communication, this paper looks into issues surrounding women's breast cancer, such as the challenges and barriers to breast cancer care, the relevant ethics and responsibilities, the right to health, breast cancer risk perception and risk communication, and the media interventions that affect Venezuelan women's perceptions and actions pertaining to this disease. In particular, it describes an action-oriented research project in Venezuela that was conducted over a four-year period of collaborative work among researchers, practitioners, NGOs, patients, journalists, and policymakers. The outcomes include positive indications on more effective interactions between physicians and patients, increasing satisfactions about issues of ethical treatment in providing healthcare services, more sufficient and responsible media coverage of breast cancer healthcare services and information, a widely supported declaration for a national response against breast cancer in Venezuela, and the creation of a code of ethics for the Venezuelan NGO that led the expansion of networking in support of women's breast cancer healthcare.

INTRODUCTION

The "right to health" in Latin America faces many challenges, which are apparent in the great health disparities present in this region. Thus, despite the World Health Organization's indication of universal health availability by the year 2000, this is not the case for Latin America (Yamin, 2000). Although Latin American women have a lower incidence rate of breast cancer than non-Hispanic women (Can-

DOI: 10.4018/978-1-5225-0549-5.ch023

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

cer mortality..., 2005, October 6), this rate is rising faster than any other female ethnic group (Buki et al., 2004). Increasing incidence rates of breast cancer among Latin American women are attributed to lifestyle changes, cultural beliefs, and lagging awareness about the disease (e.g., Andersen et al., 2004; Harvard Medical..., 2007; Schutt, Cruz & Woodford, 2008). The dilapidating effects of health disparities necessitate action from governments and human rights non-governmental organizations (NGOs) as well as media interventions in order to decrease such incongruence and address fundamental health and well-being rights.

The Venezuelan context, in particular, presents several challenges including the fact that poverty is highly prevalent and exclusion on various levels is a reality. Even though the government claims that it has reduced the overall level of poverty, indicators still show that almost 40% of Venezuelans can be considered poor and almost 12% live in extreme poverty (Ponce, 2009). A double-factor of exclusion and marginalization is being both poor and women; hence, unemployment is usually higher among women (Boza, 2004) and their income tends to be lower that the income of poor men (Orlando & Zúñiga, 2000). Exclusion is also a function of geography because some regions in the country present higher levels of birth and mortality rates, showing a clear divide between the center, the east, and the southwest (Freitez, 2003). From a health perspective, 80% of all deaths in the country are related to non-communicable diseases, while the mortality of women from 45 to 64 is mainly due to cervical and breast cancer rather than stroke or diabetes (OPS, 2011).

The 1999 Venezuelan Constitution includes three articles devoted to the right to health. These articles[1] conceive health as a "fundamental social right that must be guaranteed by the state, without discrimination of any kind, as a part of the right to life, expressing the link between health, quality of life, and collective well-being" (Feo & Siqueira, 2004, p. 372). The articles define three mechanisms for enforcing this constitutional mandate to allow for the implementation of these rights into reality: the creation of the Public National Health System, fundamental principles for the direction of the system, and the responsibility of the state to fund the system. The result of this new constitution has provided economic and social resources to populations, improving the health of Venezuelans. While such advancements are undoubtedly beneficial, further development in the realm of legislation and institutional capabilities is still required.

This paper aims to discuss issues surrounding women's breast cancer such as challenges and barriers to breast cancer care, the relevant ethics, rights, and responsibilities, breast cancer risk perception and risk communication, and the media intervention. In particular, the paper aims to describe an action-oriented research project[2] that we developed in Venezuela about women's breast cancer, which was conducted over a four-year period of collaborative work among researchers, practitioners, NGOs, patients, journalists, and policymakers.[3] It discusses four levels of ethical interactions and relationships among the stakeholders (patients, activists, physicians, public officials, journalists, scholars, and private sector representatives) in this project: 1) The basic and foundational physician-patient communication that entails information asymmetry and reveals several human, moral, and emotional dilemmas; 2) The social activism strategies as ways to empower traditionally marginalized actors (poor women and people living in remote areas) and to give voice to those who were voiceless; 3) The right to information and communication in the public sphere with a special understanding of the role of media and journalists in the portrayal of women's breast cancer, the people affected by the disease, and the need to improve access to healthcare services; and 4) The responsible exchange among patients, activists, and journalists

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

in order to have a better understanding of the media's role in shaping public perceptions, expectations, beliefs, and behaviors in relation to breast cancer, and increase the collaboration between these actors in promoting awareness and defining the public agenda.

CHALLENGES OF WOMEN'S BREAST CANCER

Understanding the most important *general*, *socioeconomic*, and *cultural* barriers are of great significance in countering the challenges of women's breast cancer. In general, early detection of breast cancer through screening (i.e., breast self examination or mammography) is the most effective method for increasing survival rates of this disease (Cazap et al., 2010). While public health campaigns can positively influence intentions to seek screening, proper actions are not commonly carried through (Catalano et al., 2003). Lacking compliance is especially prevalent among Latin American women, necessitating the identification of barriers inhibiting such behaviors to aid in the creation of strategic communication methods to mobilize positive changes. A main limitation in the control of breast cancer among Latin American women is related to disease awareness (Schwartsmann, 2001). Knowingness is interconnected with access to education, information, and healthcare, which ultimately impact women's perceived risks (Rogers, Thompson & Robinson, 2002). Lack of knowledge fuels misunderstandings, often resulting in failure to seek proper screening and treatment options (Salant & Gehlert, 2008).

Many barriers associated with unhealthy screening behaviors among Latin American women are related to fear of the disease, resulting in either excessive screening or avoidance (Bowen et al., 2004). Forbearance to seek screening is related to concerns pertaining to the implications of test results on one's family (e.g., Agurto et al., 2004; Barg & Grier, 2008; Harvey & Strahilevitz, 2009; Juraskova et al., 2007), treatment procedures (Howze, Broyden & Impara, 1992), and treatment outcomes, such as hair loss and surgical interventions (Ellingson & Buzzanell, 1999). The stress and anxiety associated with breast cancer can spark emotions of worthlessness, isolation, and loneliness, particularly due to the life threatening nature of this disease (Shaw et al., 2000).

Multiple variables interact with breast cancer screening and treatment cognizance among Latin American women; however, socioeconomic and cultural factors are the most prominent. Latin American women face many access barriers that are commonly associated with socioeconomic factors (Lantz et al., 2006). Financial constraints experienced by various sectors of society (e.g., government, institutions, and advocacy groups) impact the quality and accessibility of care for citizens (e.g., Kahn, Yang & Kahn, 2010; Luciani & Andrus, 2008).

Governmental resource deficiencies can create obstacles that hinder or completely impede citizens' access to proper care. Limited funding in developing countries often results in fewer healthcare institutions and insufficient staffing (Bechtel & Ness, 2010). A lack of healthcare practitioners can also illicit various communication barriers related to language, culture, and gender. Given the personal and sensitive nature of women's breast cancer screening and treatment, such barriers can motivate women to abstain from both seeking out and continuing screening and treatment (Montenegro & Stephens, 2006).

Socioeconomic factors, such as income and education, also impact citizens' access to proper healthcare (Herbison & Lokanc-Diluzio, 2008). Impoverished Latin American women face barriers related to proximity and transportation, as distance and inability to physically mobilize to clinics are problematic (Buki et al., 2004). Women also fear the financial burden of breast cancer; not only because it can impede a woman's ability to work, but because of high treatment and medication costs (Kapusta et al.,

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

2005). Low health literacy pertaining to breast cancer is linked to screening and treatment delinquency and avoidance or delay of discussion about symptoms with physicians (e.g., Errico & Rowden, 2006; Holt, Lee & Wright, 2008; Robles & Galanis, 2002). Due to limited health knowledge, Latin American women are commonly "misinformed about the causes of cancers, prevention of cancers, strategies for early detection of cancers, and the optimal treatments for cancers which leads to serious errors, omissions, and resultant health problems" (Kreps, 2008, p. 206). While a preliminary review of the socioeconomic barriers reveal financial support to be a prominent solution to the various issues, it is important to acknowledge the kaleidoscopic nature of such challenges. Beyond the monetary necessities, it appears that limited awareness perpetuated by multiple variables contribute to the lag in screening compliance among Latin American women.

Cultural sensitivity is considered a vital factor in encouraging Latin American women to seize ownership of their breast health. It is important to be cognizant and respectful of their cultural beliefs rather than to ignore their perceptions (Chavez et al., 1995). Due to the high value placed on family, Latin American women associate a breast cancer diagnosis with the induction of stress upon their loved ones. This concern is amplified among females due to worries regarding their ensuing incapability to continue household responsibilities and inability to protect their family from the emotional distress caused by the diagnosis (Nápoles-Springer et al., 2009). Such fears create screening barriers, which reinforce secretive behaviors and the stigmatization of cancer, further engendering breast cancer as a taboo subject (Buki et al., 2004). Despite patient fears regarding the well-being of one's family members, loved ones are also prominent sources of support for those undergoing screening and treatment, especially from other female family members (Gonzalez, Gallardo & Bastani, 2005).

Another prominent value in Latin American culture is that of trust (Borrayo, Buki & Feigal, 2005). Specifically, Hispanic women perceive physicians to be less credible and trustworthy than many other demographics, such as Caucasian or Black women, leading them to instill less faith in physicians regarding health information and suggestions (Fox & Stein, 1991). This lack of trust can motivate Latin American women to feel uncomfortable about screening and treatment, which is commonly fuelled by ideas of indecency and embarrassment regarding the exposure of their breasts or discussing breast health with a physician (McCalman, 2003). Respect for their religious beliefs is also important, especially surrounding fate and God's destiny (e.g., Buki, Salazar & Pitton, 2009; Jandorf, Bursac & Pulley, 2008). If Latin American women believe that their health outcomes are controlled by a higher power, then they may be less likely to proactively seek ways to change life outcomes, seeing them as futile (Otero-Sabogal et al., 2003).

While such challenges, barriers, and other cultural precepts create obstacles that impede proper breast cancer screening and treatment compliance, understanding and sensitivity of these norms can aid in the development of productive communication strategies that can decrease breast health delinquency among Latin American women.

ETHICS, RIGHTS, AND RESPONSIBILITIES

In the fight against women's breast cancer, issues of ethics, rights, and responsibilities are always encountered. It is, therefore, important to discuss *the right to health*, *ethics in risk communication*, and *responsibilities of public health campaigns*.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Ethical standards and principles related to the healthcare sector are complex, requiring meticulous attention and evaluation. Practitioners within the healthcare industry are responsible for their ethical conduct in care and communication with patients and their families. For instance, in addition to their standard duties, physicians are expected to enact privacy, respect, and trust to encourage autonomy among patients and increase independent and informed decision-making upon those seeking care (e.g., Davidson & Mills, 2005; Kowalski et al., 2009). Thus, health promotion strategies should be based on social values and moral concerns (Lewens, 2010).

A similar macro perspective on the study of healthcare also reveals that medical science is often far removed from the general public due to its complexity and institutionalization. This lapse can be minimized through the creation of deliberative spaces that foster public evaluation and contestations of healthcare practices. The inclusion of the public in the medical system directly pertains to issues related to healthcare ethics, as citizens must be fully informed and active within this realm (Lidskog, 2008). To leverage the decision-making power of the public and draw stronger connections between community members and those in authority, a human rights based approach to public health is deemed useful (Chopra & Ford, 2005). This method prizes community members with the right to health and development, labeling the government and public health agencies as duty bearers, with the responsibility to protect these rights and govern accordingly. Infusing the public with knowledge through effective communication strategies enhances citizen participation and productive social development.

The ethical evaluation of risks and communicating such is exceedingly complex, as it involves many different aspects and actors. Decision-makers assume moral responsibility for exposing others to risk depending on the complicity of those exposed to harm, the reasonableness of the decisions made, and how the harms and benefits are distributed (Athanassoulis & Ross, 2010). Risk decisions must be evaluated in accordance with a comparison to a virtuous ideal based on whether the agent's reasoning reflects a full and clear view of the issue. These judgments "are not irrational gut reactions, but moral intuitions and emotions that are sources of ethical wisdom that have to be taken seriously" (Roeser, 2010, p. 188). Therefore, risk communicators must strive to educate and advocate for the development of sound and productive decision-making (Bostrom, 2008).

Decision-making may be considered one of the greatest ethical concerns faced in healthcare today. Patients are bombarded by information from a plethora of sources (e.g., family, friends, independent research, and healthcare providers), and are expected to appraise various options, understand the risks involved, and enact suitable conclusions. Thus, the relationship between risk management and risk communication is innate to decision-making within healthcare due to the critical nature of such situations (Eid, 2014). Information becomes an inherently critical element in this process, as patients' understanding of accurate knowledge is necessary for reaching sound decisions (Allen, Petrisek & Laliberte, 2001).

The risk associated with decision-making is not exclusively applied to the decision-maker; it may affect other people (stakeholders), making those affected sometimes at the mercy of the decision-makers. In this, the relationship between those affected by the decision and the decision-makers becomes asymmetrical; stakeholders potentially bear more harm if faced with risks that threaten their values, identity, health, assets, and interests (Boholm, 2008). Risk also exists in the failure not to make a decision, which in many cases can cause greater harm than choosing to make one. In this, uncertainty and value judgments often coincide with risk issues and decision-making, generating varied interpretations (Bakir, 2010).

Closely related to risk in decision-making is uncertainty, which refers to not knowing the precise outcome of the choice enacted. A useful classic model in facing uncertainty and risk when making decisions is the cost-benefit analysis (CBA), which is a rational model. The CBA requires one to weigh the

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

potential costs and benefits of the option and to compare these costs and benefits with all other options or alternatives available (Clarke, 2010). However, it is rare that an individual can ever be fully rational when making decisions, especially when they involve large consequences pertaining to one's family or health. These circumstances may render the CBA model for decision-making less effective than claimed by its proponents. However, while the CBA may not always seem relevant in patients' decision making, it is believed that through proper communication with physicians and loved ones and the provision of information patients can reach rational conclusions.

To aid physicians and patients through the subjective path of decision-making, a prescriptive model is proposed that encourages healthcare practitioners to follow a broad ethical framework (e.g., professional code of ethics) to guide communicative actions to patients about decisions (Addington & Wegescheide-Harris, 1995). This allows practitioners to evaluate actions, judgments, and theories, to illuminate the ethical presuppositions founding patients' decisions, and to understand how irrational factors impact conclusion enactment.

Public health campaigns can be used for many purposes; however, there is a barrage of responsibilities and ethical concerns related to these communicative processes. Government and healthcare practitioners face many dilemmas regarding their conduct in public health communication campaigns, necessitating a detailed analysis of such to guide the development of sound and productive messaging. While some ethical concerns can be based strictly on measuring how an audience may be immediately impacted by a message, there are other problems related to public campaigns. For example, while some campaigns may pose obvious ethical issues, harm can also fall upon society inadvertently when campaigns seek to reduce behaviors, which may deprive individuals of pleasure or of significant cultural activities (Guttman, 1997). Additionally, it is also important to understand how messages about health dictate ideals and morals. For instance, some campaigns that advocate certain behavior may (inadvertently or intentionally) tag those who comply with the behavior as "good" and those who do not as "bad".

Beyond the central focus on overarching values for healthcare, it is especially important to ensure that healthcare communication strategies take into account contextual and cultural factors, such as moral and religious values. Such principles form the foundation of social groups, villages, cities, and nations alike and may affect patient, physician, and authority perceptions and evaluations of healthcare. In a Latin American context, many cultural values need to be considered when developing public health campaigns. For instance, the values of collectivism, family, religion, trust, respect, and interpersonal relationships are deemed significant (Borrayo, 2004). Therefore, health messages must be sensitive to these principles and include content that interacts with or reinforces such values.

RISK: PERCEPTION AND COMMUNICATION

Risk perception can be understood as a "value-dependent process" (Etkin & Ho, 2007, p. 627), influenced by inherent social values that determine which risks are selected for assessment and what tools are to be used in their measurement (Mairal, 2008). Risk assessment also involves emotional and intuitive responses, which impact the deliberation process of how to deal with, or mitigate situations (Roeser, 2010). There are generally two separate ways in which risk is perceived: the first is an affective process, referred to as "the primary evaluative process" that involves a typically instantaneous gut, intuitive reaction to information; and the "secondary evaluative process", which is deliberative, usually taking place after the initial reaction to information and involves the thinking-it-through period (Visschers et al., 2008,

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

p. 207). While both reactions are to be expected, often the primary, intuitive reaction tends to dominate over deliberation, which can lead to misperceptions of risk. Whether overestimating, underestimating, or simply not understanding one's risk, the results can be detrimental and lead to action or inaction, which may negatively affect one's health.

Research highlights a vast array of risk perceptions about women's breast cancer, which interacts with lagging breast cancer screening compliance. For instance, patients greatly overrate their risk of developing or dying from the disease (Dubé, O'Donnell & Novack, 2000). Such perceptions are often related to women's misunderstandings of information related to breast health and screening (Wonderlick & Fine, 1997). With so much information to take in, coupled with the stigma and fear surrounding breast cancer, it is unsurprising that misperceptions about one's own risk are common. Common misperceptions about cancer among Latin American women (e.g., cancer is a death sentence, cancer cannot be cured, and cancer may be contagious) illuminate factors that may contribute to risk misunderstandings (Granda-Cameron, 1999). Latin American women are more fatalistic in their perceptions of breast cancer than other ethnic groups and tend to overestimate their risk for getting the disease (Mishra et al., 1998). They tend to believe that it is a symptomatic disease, causing them to greatly underestimate their risk for getting breast cancer, and consequently, they were much less likely to seek screening (Borrayo, Buki & Feigal, 2005). It is, therefore, suggested that accurate risk comprehension and perceived susceptibility to breast cancer is associated with increased screening behavior among women (Williams, Clarke & Savage, 2002). Providing information to women regarding the negative consequences of failing to practice breast screening may increase knowledge of risk and also motivate women to engage in proper screening practices (Meyerowitz & Chaiken, 1987).

Risk communication is considered significant due to the importance of disseminating messages regarding risk to properly inform audiences on the threat of women's breast cancer and potentially mobilize and activate populations to engage in healthy behaviors. Successful risk communication "requires establishing effective communication channels and developing mutually respectful relationships between patients and providers" (Nekhlyudov & Partridge, 2003, p. 236). Thus, techniques should avoid paternalistic and persuasive approaches and focus on fostering open and equal communication when discussing risk. Effective health risk communication is dependent on understanding the information needs of recipients and can directly impact peoples' decision-making in relation to health behaviors, such as mammography practices.

Various studies identify risk communication as a concept that is socially imbedded in different interested communities, such as academics, interest groups, governments, and the public (e.g., Bergmans, 2008; Hansson, 2010; Stoffle & Minnis, 2008; Strydom, 2008). Psychological, cultural, social, and institutional processes are intervening factors that tend to enhance an individual's perception of risk, subsequently influencing risk behavior. Cultural constraints in Latin America include cultural norms of modesty, distrust, and limited familiarity with the healthcare system (Kreps & Sivaram, 2008). Trust, for example, is essential for successful risk communication (e.g., Blennow, 2008; Löfstedt, 2010); it results from transparent and honest communication, which circumvents uncertainty among the public.

Effective risk communication requires collective discussions among all concerned parties. Risk communication can be improved through the integration of potential stakeholders, along with officials, in the definition of the risk problem, aiding in the unification of differing opinions and views about risk and allowing for the discussion of issues associated with the risk from varying perspectives (Brunet & Houbaert, 2007). Despite the introduction of patient-centre care paradigms, physicians still often lack proper training on best practices for effective risk communication (Hanoch, 2004). For example, in

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

communication with female patients about their mammogram results, it is more effective and preferred by patients if communication occurs in person, rather than over the phone or other means, because it allows physicians and counselors to question patients on their perception of their risk in order to avoid misconceptions (e.g., Carney et al., 2009; Quillin et al., 2008). Latin American women are more likely to comply with healthy screening behaviors and exhibit accurate awareness of risk when they are provided with information that they can understand; for instance, through counseling and guided discussions, their perceptions of risk demonstrate increased accuracy and their fears of cancer are decreased (Bowen et al., 2004). Accuracy of risk assessment among Latin American women increases when physicians provide them with social comparison information of other women with higher risk of breast cancer (Dillard et al., 2006). Being prepared to discuss risk to family members is also vital, as one of the greatest fears Latin American women feel is how a diagnosis may negatively impact their family members (e.g., Luker et al., 1995; Macdonald et al., 2008). This fear may act as an additional barrier to women seeking screening for breast cancer. Thus, breast cancer risk should be framed in a way that alleviates women's anxiety over the health of their family members.

MEDIA INTERVENTION

Breast cancer has attracted notable media coverage, which is likely attributed to the growing prevalence of this disease and to society's heightened fixation on risks and fear regarding health issues (Cookson, 2010). Exposure to media regarding women's breast cancer screening practices is shown to encourage effective uptake of this behavior among women (Wang et al., 2008). The use of storytelling in the media is employed as a method of conveying the experiences of this disease in society and how it impacts the lives of women and their surroundings (Kreuter et al., 2007). The vast majority of stories about breast cancer provide citations of specific individuals, typically experts or women experiencing the disease; a process of personification through narratives that allows the audience to connect to breast cancer on a unique level, characterizing its realities and reifying survivors (Atkin et al., 2008). However, debates over the representations of the fusion of women and breast cancer in the media face criticism. Breast cancer is a disease that impacts women globally; women of all different demographics face the burden of this disease in some capacity. Despite this, mainstream media representations tend to leave out the dynamic nature of the people who are affected by breast cancer—often failing to incorporate "the complexities of age and beauty as they pertain to specific groups of women" (Cartwright, 1998, p. 131).

Agenda-setting theory posits that the media play a pivotal role in shaping public thought and discussion on breast cancer screening behavior compliance (Eid & Dakroury, 2012). It may be possible to promote changes in health behaviors with the help of media strategizing, as journalists and media practitioners have the ability to impact public health in a positive manner due to the power of agenda setting, potentially resulting in opinion changes (Dakroury & Eid, 2012; Jones, Denham & Springston, 2006). For example, a study about the impact of exposure to newspaper coverage on breast cancer screening practices among female college students and their mothers revealed that "the quantity of mammogram screening coverage in newspapers contributes to mammography utilization" (Atkin et al., 2008, p. 6). Media coverage of public figures and their experiences with breast cancer and widely publicized events that draw attention to this disease and the importance of screening can be an opportunity to persuade women to engage in health behaviors, such as breast self-examination and mammography (Fink et al., 1978).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Activism and advocacy efforts have also attracted significant media coverage due to growing community involvement in breast cancer awareness (Cartwright, 1998). Communication "through the local media can be an invaluable resource for mobilizing community support and activity" (Figueroa et al., 2002, p. 10). Agencies and organizations that specialize in cancer research or awareness often work to establish relationships with media outlets with the intention to promote knowledge on the nature, detection, and treatment of breast cancer (Catalano et al., 2003). However, while the media are said to be an important health communication tool in the dissemination of information about women's breast cancer, it is also important to acknowledge the limitations involved in holding expectations of this transmission vehicle's standards. The media may fail to provide audiences with sufficient information, resulting in distorted understandings or conceptions of the disease. Thus, while media coverage on this topic is inevitable, it is important to note that not all is considered helpful in breast cancer health communication ventures. For example, the news media rarely provide sufficient information regarding pertinent topics related to breast cancer, such as prevention methods and encouragement of healthy dietary practices (Atkin et al., 2008). Also, while the Internet is considered a valuable tool with regards to communicating information about breast cancer as women can seek knowledge about the disease or interact in virtual communities to obtain information or support, there are many limitations associated with information acquisition via this medium, such as the lack of trust regarding the credibility and accuracy of the information (e.g., Bernhardt et al, 2002; Elliott, 2007; Lee et al., 2008; Winzelberg et al., 2003). Thus, in exploring the analytical agenda of media involvement in health communication associated with breast cancer, it becomes clear that further investigation on the role of the media is required.

METHOD AND OUTCOMES: INITIATIVE FOR ACTION IN VENEZUELA

Action-oriented research promotes a critical consciousness that exhibits itself in political and practical action to make some type of change. Its agenda mainly includes "producing research that can address practical concerns" (Small, 1995, p. 942), aiming to "have some type of change" (Small & Uttal, 2005, p. 938). It helps to: 1) increase the closeness between the day-to-day problems encountered by practitioners in specific settings and the theories used to explain and resolve the problem; and 2) assist practitioners in lifting their veil of clouded understandings to better understand fundamental problems by raising their collective consciousness (Berg, 2009; Eid, 2011). As "a methodological approach for doing collaborative research with practitioners and community partners that can inform practice, programs, community development, and policy while contributing to the scientific knowledge base" (Small & Uttal, 2005, p. 936), action-oriented research is a collective, self-reflective, and participatory enquiry undertaken by participants (stakeholders) in social interactions to improve some condition in which they are all involved.

Action-oriented researchers "value collaboration with nonresearcher participants" in that while they bring "to the research process theoretical knowledge, experience, and the skills of conducting social science research, the participant collaborators bring practical knowledge and experience about the situations that are being studied" (Small, 1995, p. 942). They "believe that the process of conducting research should be empowering for research participants" (Esterberg, 2002, p. 141). Aiming to "generate knowledge that can be used to address practical concerns of local communities, organizations, and groups and incorporate local understandings of specific practices and issues" (Small & Uttal, 2005, p. 938), action-oriented researchers and community partners "combine their different kinds of knowledge and skills to produce insightful and usable findings" (Small & Uttal, 2005, p. 938).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

We developed an action-oriented research project in Venezuela in collaboration with SenosAyuda, an NGO that supports women with breast cancer and advocates for the early detection of the disease and better access to healthcare services, as well as a group of physicians, patients and survivors, journalists, and policymakers. Our research has been guided by the "ecological perspective" (Street, 2003) that proposes an integration of the different levels in health communication (interpersonal, organizational, and social), and highlighting the potential emancipation role of health communication according to the pioneer of the field, Luis Ramiro Beltrán (1994). This project aimed at helping SenosAyuda in the creation of a national network of breast cancer volunteer organizations while facilitating the acquisition of communication and advocacy skills and capabilities of breast cancer survivors and community activists.

The project was implemented in a very particular political, social, and cultural environment. Since 1999, when Hugo Chavez became president after winning a democratic election, Venezuela has lived under a highly and virulent social and political polarization. The divide between the Chavez's followers and the opposition has translated into political violence, attacks to the freedom of speech, and a discourse of hate from both sides. The media have played an important role in the dynamics of polarization. The Chavez's regime has contributed to this polarization by promoting what the government calls the "new media hegemony" through the control of a vast network of radio and TV stations, newspapers, and Internet websites that are part of a huge propaganda machine (Bisbal, 2009). On the other hand, the Chavez' regime has put a strong emphasis on social programs in order to help the poor and expand the access to services, including access to healthcare, specially through its flag program *Misión Barrio Adentro* (Inside the Neighborhood Mission) with the help of the government of Cuba (Briggs & Mantini-Briggs, 2009). The overall performance of the *Misión* is far from the positive results announced by the government; this program has even contributed to increase the historical fragmentation of the Venezuelan healthcare system with a negative impact on the overall financing of the public sector (e.g., Díaz Polanco, 2006; Nahon-Serfaty, 2005).

The various social challenges in Venezuela include insufficient and irresponsible media coverage of women's breast cancer and the relevant healthcare services and information, general resistance from the public to learn about women's breast cancer, shortage in physicians' availability, lack of training of spokespeople on behalf of patients, survivors, and foundations, shortages in the logistics of organizing public events, absence or reluctance of volunteers and advocates, deficiencies in activity evaluation and feedback measures, and incompetence in the physician-patient interactive relationship.

Culturally, *machista* chauvinist prejudices still dominate the public debate about breast cancer, as shown for example by the controversy prompted by soccer fans protesting because their team decided to wear pink shirts during the breast cancer month in October 2012. According to a news report, the suspension of the game between Deportivo Táchira and Atletico Venezuela was ordered by representatives of the Venezuelan Football Federation after fans refused to leave the field for 40 minutes "to protest the pink shirts being worn in support of the fight against breast cancer, led by the Fundación SenosAyuda" (Bailey, 2012, October 29).

As a way to fill the void left by the public sector in the provision of healthcare, especially to the poor populations, the civil society organizations emerged in Venezuela. *Sociedad Anticancerosa de Venezuela* (the Cancer Society of Venezuela) was the oldest organization, created in 1948, to lead the efforts to educate and provide medical care to those lacking financial means or insurance coverage. Recently many NGOs working specifically in breast cancer have been created: SenoSalud, Funcamama, Fuccam, Famac, Fundamama, Fundaseno, Senosestima, Unicam, and SenosAyuda, among others.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

To address the problems and deficiencies that women are facing with breast cancer issues and to promote nationwide long-term action programs, SenosAyuda was created in 2007. This NGO initiated the Organizational Capacity Building Program (OCBP) directed to collaborate with other social development organizations involved in the fight against women's breast cancer in Venezuela. The OCBP objectives were: 1) to contribute to the development of such organizations in three fundamental phases of their evolution: to be born and established, to progress, and to be consolidated or redefined; and 2) to educate the leadership hands-on learning methodologies and provide the social development organizations with the managerial tools that allow them to manage their associations efficiently.

Within the context of the OCBP initiative, our project aimed to achieve four main goals: 1) to develop and enhance the communication capabilities of activists in order to increase their impact as spokespersons of their cause; 2) to identify needs and gaps in the communication between different stakeholders (i.e., physicians, patients, activists, journalists, and public officials); 3) to promote open discussions and debates about the ethical issues relevant to women's breast cancer in order to achieve more transparent, symmetrical, and empowering engagement and action among stakeholders; and 4) to create networking opportunities in order to expand the impact of social civil organizations and activists in the public discourse about breast cancer and the need for a national policy.

Through four annual collective workshops in Caracas and continuous national events and activities during the four-year period (2009-2012), collaborators of this action-oriented research project have engaged in discussions. The workshops essentially helped participants to be spokespeople for their organizations in the media and to understand how to use communication as a tool to plan and achieve the given activities. The workshops provided skills of ethically and responsibly composing and communicating messages about breast health that can help promote changes in attitude and behavior. They helped with ethical media relations and how to define the goals of interviews and to target the audiences and select the strategies to reach them. The workshops also included discussions on the situation of patients living in regions with no access to healthcare services, the confusion caused by the debate about breast implants, and the need to secure humane and dignified palliative care.

During the workshops, discussions have highlighted the media influence in the fight against women's breast cancer. It has been noted that although patients generally prefer screening information supplied by primary care providers, their providers actually play a limited role; most information come from media sources. Media campaigns impact awareness of mammography as a screening tool, knowledge of mammography guidelines, intention to be screened, and motivation to seek first time and subsequent mammograms. Criticisms of media practices include that news coverage of preventive actions focuses mainly on what women can do at the individual level rather than portraying actions at the family or societal level. Also, news media present almost no information to educate parents about how to protect their daughters from environmental risks. Useful and successful media practices were suggested, such as that message strategies that include testimonials from real-life survivors of breast cancer significantly affect participants' willingness to seek early detection for breast cancer. The workshops' discussions have emphasized fundamental ethical principles that are necessary for the media to abide by when covering issues surrounding women's breast cancer. These principles include independence, accuracy, integrity, truth, fairness, and serving the public interest (Eid, 2008). Moreover, protecting the patients' privacy, showing respect, and correcting mistakes were discussed.

A wide range of events and activities were conducted by collaborators in order to promote breast cancer health awareness. October 19th was the day dedicated to promoting breast health. In particular, "The Pink Walk" was an important event where breast cancer survivors and other activists from the

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

community could walk the streets, raising awareness. In addition, posters and banners were put up in public places such as bus stops and taxi and moto-taxi stops. Another event was a fashion show where breast health was promoted at both the regional and national levels (television coverage by the station TELEVEN). Other events included tennis tournaments to raise funds and support the foundation, funding campaigns, and the design of a project to strengthen the development of breast health education.

Activities include the production of materials, such as "If you read me, you take care of yourself and you touch yourself", which was distributed to the public. Also, the "Pink Recipe," which discusses the four "to-dos" for breast health: self-exam, breast echography, mammogram, and consultation with a physician. Other materials were newsletters, letters asking for contributions and thanking sponsors and volunteers, and website updates and maintenance. Essentially, the most significant activities involved getting the messages of breast health out to the public, the media, and even physicians. The public was reached through direct talks (such as the ABC's of breast cancer—talks about the organization, activities, and early detection and prevention), distribution of pamphlets, information, and pink ribbons. For example, the talks with the Mission Mothers of Barrio helped reach women in poorer communities. The creation of clubs called, "Prevention and Early Diagnostic of Breast Cancer", involved different talks about breast health, the creation of billboards and newsletters, and even theatre presentations. The media were beneficial through conducting interviews to help get the messages out, and the physicians were of assistance in providing information as well as being spokespeople to help further convey the importance of breast health.

A major outcome of our action-oriented research project has been the expansion of the networking base of SenosAyuda with different stakeholders including other NGOs from different regions, physicians, scholars, journalists, and public officials from state and municipal governments.[4] For example, a direct outcome celebrated in May 2011 after the third workshop of our project was the declaration "*For a Consensus Vision for the formulation and implementation of a National Response against Breast Cancer in Venezuela*," which was subscribed by an initial group of 20 participants. A year after, in May 2012, the document has been signed by 101 organizations, including 8 breast cancer advocacy organizations, 6 major medical associations, 4 human rights NGOs, and important social actors, such as the Red Cross and the National College of Journalists. The declaration has also contributed to set additional goals of networking and consensus building for SenosAyuda, the organization that is leading the initiative, including the collection of 10,000 signatures by the end of 2013 in order to support a National Response against Breast Cancer in Venezuela, the launch of a national media campaign to raise public awareness about the need of a consensus, and the celebration of public forum all over the country to discuss the principles and objectives that should guide a national policy for the disease. Moreover, the declaration has contributed to highlighting some of the major ethical issues in breast cancer information and care, particularly the need to communicate messages associated with the solutions available to the public and adapted to their needs and socio-cultural realities, the role of the state in securing financial protection to every person affected by breast cancer through public and private healthcare insurances and funding, and the provision of palliative care as part of human, moral, and ethical duties towards people suffering from breast cancer.

Another major outcome of our action-oriented research project has been the creation of a code of ethics for SenosAyuda. In light of the NGO's mission, this code aims to emphasize a commitment to: 1) maintaining social responsibility; 2) upholding human rights; 3) increasing breast cancer awareness; 4) ensuring transparency; and 5) maintaining corporate responsibility. These major ethical principles have been emphasized through detailed relevant ethical standards and guidelines. To maintain social

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

responsibility, SenosAyuda serves the communities within which it operates with a focus on helping those affected by breast cancer, upholds the reputation of communities it represents, complies with all government regulations, adopts a positive and proactive role in the community, and takes an active role in the creation and moderation of breast cancer care services. To uphold human rights, SenosAyuda respects and acknowledges the dignity and worth of all individuals involved, respects cultural, individual, and role differences among people, upholds the individual's rights to privacy and maintains the confidentiality and security of their information, recognizes the individual's innate dignity as a human being and respects their autonomy and freedom of choice, and ensures that donors, beneficiaries, members, and others are free from undue pressure arising from personal relationships, financial or gift incentive or disincentive, and other forms of coercion. To increase breast cancer awareness, SenosAyuda's education and information materials and programs are created for, aimed at, and made accessible to those affected by breast cancer in the community. SenosAyuda retains transparency (Eid & Nahon-Serfaty, 2008) in all of its dealings with the government, the public, donors, partners, beneficiaries, and other interested parties, presents information in a fair, accurate, and unbiased manner, is obliged to disclose conflicts of interest whenever they appear, and provides accurate and timely information on the procurement and management of donations received. Finally, to maintain corporate responsibility, SenosAyuda uses all funds solely for the purpose of continuing its mission, actively seeks out new donors and opportunities to raise funds for its cause, and uses donations in accordance with the commitments it has made to contributors.

CONCLUSION

To conclude, this paper describes an action-oriented research project that we developed in Venezuela to help women reduce the risk of breast cancer by discussing several levels of ethical interactions and relationships among the stakeholders (patients, activists, physicians, public officials, journalists, scholars, and private sector representatives) which are currently social-cultural obsticales towards this goal.

The key to reducing the risk is to foster positive health attitudes and behaviors among women to understand what is effective in removing general, cultural, and socioeconomic barriers to seeking screening or treatment. Important barriers for Latin American women related to screening are anxiety about the procedure, cost, embarrassment, fear of the pain of mammography, and fear of being diagnosed with breast cancer, which may be perceived potentially as a death sentence or serious issue for their family. Also, due to the high value placed on modesty among Latin American women, and the fact that with time and experience they become increasingly comfortable touching their breasts for self-examination, they may benefit from physicians and nurses showing them how to conduct a breast self-exam. Therefore, in the promotional efforts for breast self-examination, it is essential to understand and be sensitive to these cultural beliefs and feelings and not simply attempt to educate these beliefs away. The physician-patient discussion of breast cancer and screening is a strong factor in the likelihood of women's adherence to breast exams (Smith et al., 2009). In this physician-patient relationship as well as many other interactions and practices relevant to women's breast cancer, ethical principles, rights, and responsibilities are fundamental to reach positive outcomes.

A perception of risk is a prominent driving force to seek breast cancer screenings and practice healthier behaviors (Hubbell, 2006). In creating strategic risk communication messages that can encourage breast cancer early detection for women, it is essential to frame messages in a manner that address cultural constraints such as women's modesty, religion, distrust of physicians or the healthcare system in general,

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

and low socioeconomic status, which may all prevent women from seeking breast cancer screening (e.g., Kreps, 2008; Kreps & Sivaram, 2008). Messages for Latin American audiences should also recognize the heterogeneous and diverse nature of this population and demonstrate cognizance of communities with low health literacy rates (Buki, Salazar & Pitton, 2009).

Despite the fact that governmental efforts strive to improve healthcare accessibility and promote equality, deficiencies and disparities persist in the status of breast cancer among Venezuelan women. This issue is of specific importance, as lagging health literacy about the disease and growing incidence rates can result in increased mortality. In the fight against women's breast cancer in Venezuela, it is suggested that public health campaigns work to raise awareness on the value of screening behaviors and government strives to facilitate increased accessibility to public prevention and treatment facilities. This goal requires the acknowledgement of the barriers that impede Venezuelan women from complying with productive breast health practices and an assessment of their risk perception.

It is possible to develop effective risk communication strategies through media interventions that can influence persuasive and useful public health campaigns regarding women's breast cancer. The media coverage of this disease is specifically useful, as health beliefs and ideas are intrinsically linked to media depictions. As a result of the workshops, events, and activities of our action-oriented research project in Venezuela, there has been positive indications on more effective interactions between physicians and patients, increasing satisfactions about issues of ethical treatment in providing healthcare services, and more sufficient and responsible media coverage of breast cancer healthcare services and information.

REFERENCES

Addington, T., & Wegescheide-Harris, J. (1995). Ethics and communication with the terminally ill. *Health Communication*, *7*(3), 267–281. doi:10.1207/s15327027hc0703_5 PMID:11657458

Agurto, I., Bishop, A., Sanchez, G., Betancourt, Z., & Robles, S. (2004). Perceived barriers and benefits to cervical cancer screening in Latin America. *Preventive Medicine*, *39*(1), 91–98. doi:10.1016/j.ypmed.2004.03.040 PMID:15207990

Allen, S. M., Petrisek, A. C., & Laliberte, L. L. (2001). Problems in doctor-patient communication: The case of younger women with breast cancer. *Critical Public Health*, *11*(1), 38–58. doi:10.1080/09581590010028246

Andersen, M. R., Ankerst, D. P., Bowen, D. J., McGregor, B. A., McTiernan, A., & Yasui, Y. (2004). Optimism, perceived risk of breast cancer, and cancer worry among a community-based sample of women. *Health Psychology*, *23*(4), 339–344. doi:10.1037/0278-6133.23.4.339 PMID:15264969

Athanassoulis, N., & Ross, A. (2010). A virtue ethical account of making decisions about risk. *Journal of Risk Research*, *13*(2), 217–230. doi:10.1080/13669870903126309

Atkin, C. K., Smith, S. W., McFeters, C., & Ferguson, V. (2008). A comprehensive analysis of breast cancer news coverage in leading media outlets focusing on environmental risks and prevention. *Journal of Health Communication*, *13*(1), 3–19. doi:10.1080/10810730701806912 PMID:18307133

Bailey, R. (2012, October 29). Deportivo Tachira match abandoned after fans riot over pink Breast Cancer Awareness kit. *Yahoo News*. Retrieved November 11, 2012, from http://sports.yahoo.com/blogs/soccer-dirty-tackle/deportivo-tachira-match-abandoned-fans-riot-over-pink-181919490--sow.html

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Bakir, V. (2010). Media and risk: Old and new research directions. *Journal of Risk Research*, *13*(1), 5–18. doi:10.1080/13669870903135953

Barg, F. K., & Grier, S. A. (2008). Enhancing breast cancer communications: A cultural models approach. *International Journal of Research in Marketing*, *25*(4), 335–342. doi:10.1016/j.ijresmar.2008.07.003

Bechtel, C., & Ness, D. L. (2010). If you build it, will they come? Designing truly patient-centered healthcare. *Health Affairs*, *29*(5), 914–920. doi:10.1377/hlthaff.2010.0305 PMID:20439880

Beltrán, L. R. (1994). La salud y la comunicación en Latinoamérica: Políticas, estrategias y planes. In *UNESCO - OPS, Por una política de comunicación para la promoción de la salud en América Latina* (pp. 29–90). Quito, Ecuador: OPS.

Berg, B. L. (2009). *Qualitative research methods for the social sciences*. Boston, MA: Allyn & Bacon.

Bergmans, A. (2008). Meaningful communication among experts and affected citizens on risk: Challenge or impossibility? *Journal of Risk Research*, *11*(1-2), 175–193. doi:10.1080/13669870701797301

Bernhardt, J. M., Lariscy, R. A. W., Parrott, R. L., Silk, K. J., & Felter, E. M. (2002). Perceived barriers to Internet-based health communication on human genetics. *Journal of Health Communication*, *7*(4), 325–340. doi:10.1080/10810730290088166 PMID:12356290

Bisbal, M. (Ed.). (2009). *Hegemonía y control comunicacional*. Caracas, Venezuela: Alfa & UCAB.

Blennow, K. (2008). Risk management in Swedish forestry – Policy formation and fulfilment of goals. *Journal of Risk Research*, *11*(1-2), 237–254. doi:10.1080/13669870801939415

Boholm, A. (2008). The public meeting as a theater of dissent: Risk and hazard in land use and environmental planning. *Journal of Risk Research*, *11*(1-2), 119–140. doi:10.1080/13669870701633852

Borrayo, E. A. (2004). Where's Maria? A video to increase awareness about breast cancer and mammography screening among low-literacy Latinas. *Preventive Medicine*, *39*(1), 99–110. doi:10.1016/j.ypmed.2004.03.024 PMID:15207991

Borrayo, E. A., Buki, L. P., & Feigal, B. M. (2005). Breast cancer detection among older Latinas: Is it worth the risk? *Qualitative Health Research*, *15*(9), 1244–1263. doi:10.1177/1049732305281337 PMID:16204403

Bostrom, A. (2008). Lead is like mercury: Risk comparisons, analogies and mental models. *Journal of Risk Research*, *11*(1-2), 99–117. doi:10.1080/13669870701602956

Bowen, D. J., Burke, W., McTiernan, A., Yasui, Y., & Andersen, M. R. (2004). Breast cancer risk counseling improves women's functioning. *Patient Education and Counseling*, *53*(1), 79–86. doi:10.1016/S0738-3991(03)00122-8 PMID:15062908

Boza, M. E. (2004). *Desempleo en primera persona*. Caracas, Venezuela: IESA.

Briggs, C. L., & Mantini-Briggs, C. (2009). Confronting health disparities: Latin American social medicine in Venezuela. *American Journal of Public Health*, *99*(3), 549–555. doi:10.2105/AJPH.2007.129130 PMID:19150916

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Brunet, S., & Houbaert, P. (2007). Involving stakeholders: The Belgian fowl pest crisis. *Journal of Risk Research*, *10*(5), 643–660. doi:10.1080/13669870701281454

Buki, L. P., Borrayo, E. A., Feigal, B. M., & Carrillo, I. Y. (2004). Are all Latinas the same? Perceived breast cancer screening barriers and facilitative conditions. *Psychology of Women Quarterly*, *28*(4), 400–411. doi:10.1111/j.1471-6402.2004.00157.x

Buki, L. P., Salazar, S. I., & Pitton, V. O. (2009). Design elements for the development of cancer education print materials for a Latina/o audience. *Health Promotion Practice*, *10*(4), 564–572. doi:10.1177/1524839908320359 PMID:19098265

Cancer mortality: Researchers compare Latin American cancer mortality to other countries. (2005, October 6). *Women's Health Weekly*, p. 65.

Carney, P. A., Kettler, M., Cook, A. J., Geller, B. M., Karliner, L., Miglioretti, D. L., & Elmore, J. G. et al. (2009). An assessment of the likelihood, frequency, and content of verbal communication between radiologists and women receiving screening and diagnostic mammography. *Academic Radiology*, *16*(9), 1056–1063. doi:10.1016/j.acra.2009.02.023 PMID:19442539

Cartwright, L. (1998). Community and the public body in breast cancer media activism. *Cultural Studies*, *12*(2), 117–138. doi:10.1080/095023898335500

Catalano, R., Winett, L., Wallack, L., & Satariano, W. (2003). Evaluating a campaign to detect early stage breast tumors in the United States. *European Journal of Epidemiology*, *18*(6), 545–550. doi:10.1023/A:1024620617272 PMID:12908720

Cazap, E., Buzaid, A., Garbino, C., de la Garza, J., Orlandi, F., Schwartsmann, G., & Breitbart, G. et al. (2010). Breast cancer in Latin America: Experts perceptions compared with medical care standards. *The Breast*, *19*(1), 50–54. doi:10.1016/j.breast.2009.10.011 PMID:19945878

Chavez, L. R., Hubbell, A., McMullin, J. M., Martinez, R. G., & Mishra, S. I. (1995). Structure and meaning in models of breast and cervical cancer risk factors: A comparison of perceptions among Latinas, Anglo women, and physicians. *Medical Anthropology Quarterly*, *9*(1), 40–74. doi:10.1525/maq.1995.9.1.02a00030 PMID:7697550

Chopra, M., & Ford, N. (2005). Scaling up health promotion interventions in the era of HIV/AIDS: Challenges for a rights based approach. *Health Promotion International*, *20*(4), 383–390. doi:10.1093/heapro/dai018 PMID:16061497

Clarke, S. (2010). Cognitive bias and the precautionary principle: What's wrong with the core argument in Sunstein's Laws of Fear and a way to fix it. *Journal of Risk Research*, *13*(2), 163–174. doi:10.1080/13669870903126200

Cookson, C. (2010). How journalists perceive risk communication – risk communication guidelines for Europe: A modest proposition – comments to R. Lofstedt. *Journal of Risk Research*, *13*(1), 111–114. doi:10.1080/13669870903126150

Dakroury, A., & Eid, M. (2012). Introduction to communication and media studies. In M. Eid & A. Dakroury (Eds.), *Basics in communication and media studies* (pp. 3–31). Boston, MA: Pearson.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Davidson, R., & Mills, M. E. (2005). Cancer patients' satisfaction with communication, information and quality of care in a UK region. *European Journal of Cancer Care*, *14*(1), 83–90. doi:10.1111/j.1365-2354.2005.00530.x PMID:15698390

Díaz Polanco, J. (2006). Salud y Hegemonía Regional: Las relaciones Venezuela-Cuba, 1999-2006. *Foreign Affairs en Español*, *6*(4), 10–12.

Dillard, A. J., McCaul, K. D., Kelso, P. D., & Klein, W. M. P. (2006). Resisting good news: Reactions to breast cancer risk communication. *Health Communication*, *19*(2), 115–123. doi:10.1207/s15327027hc1902_3 PMID:16548702

Dubé, C. E., O'Donnell, J. F., & Novack, D. H. (2000). Communication skills for preventive interventions. *Academic Medicine*, *75*(7), 43–54. PMID:10926040

Eid, M. (2008). *Interweavement: International media ethics and rational decision-making*. Boston, MA: Pearson.

Eid, M. (2011). Introduction to communication research. In M. Eid (Ed.), *Research methods in communication* (pp. 3–14). Boston, MA: Pearson.

Eid, M. (2014). Editorial: Risk and crisis communication. Global Media Journal -- Canadian Edition, 7(1), 1-3.

Eid, M., & Dakroury, A. (Eds.). (2012). *Basics in communication and media studies*. Boston, MA: Pearson.

Eid, M., & Nahon-Serfaty, I. (2008). Transparency in direct-to-consumer pharmaceutical marketing: Crucial ethical standards. *IMC Review: Journal of Integrated Marketing Communications*, *7*(1), 6–14.

Ellingson, L. L., & Buzzanell, P. M. (1999). Listening to women's narratives of breast cancer treatment: A feminist approach to patient satisfaction with physician-patient communication. *Health Communication*, *11*(2), 153–183. doi:10.1207/s15327027hc1102_3 PMID:16370974

Elliott, C. (2007). Pink! Community, contestation, and the colour of breast cancer. *Canadian Journal of Communication*, *32*(3), 521–536.

Errico, K. M., & Rowden, D. (2006). Experiences of Breast Cancer survivor-advocates and advocates in countries with limited resources: A shared journey in breast cancer advocacy. *The Breast Journal*, *12*(1), S111–S116. doi:10.1111/j.1075-122X.2006.00208.x PMID:16430394

Esterberg, K. G. (2002). *Qualitative methods in social research*. Boston, MA: McGraw-Hill.

Etkin, D., & Ho, E. (2007). Climate change: Perceptions and discourses of risk. *Journal of Risk Research*, *10*(5), 623–641. doi:10.1080/13669870701281462

Feo, O., & Siqueira, C. E. (2004). An alternative to the Neoliberal model of health: The case of Venezuela. *International Journal of Health Services*, *34*(2), 365–375. doi:10.2190/NHFK-GHQW-AQLA-YC5D PMID:15242164

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figueroa, M.E., Kincaid, L.D., Rani, M., Lewis, G. & Gray-Felder, D. (2002). *Communication for social change: An integrated model for measuring the process and its outcomes*. Baltimore: Health Communication Partnership based at the John Hopkins Bloomberg School of Public Health/Center for Communication Programs and Rockefeller Foundation as part of their Communication for Social Change Grantmaking Strategy.

Fink, R., Roeser, R., Venet, W., Strax, P., Venet, L., & Lacher, M. (1978). Effects of news events on response to a breast cancer screening program. *Public Health Reports*, *93*(4), 318–327. PMID:684143

Fox, S. A., & Stein, J. A. (1991). The effect of physician-patient communication mammography utilization by different ethnic groups. *Medical Care*, *29*(11), 1065–1082. doi:10.1097/00005650-199111000-00001 PMID:1943268

Freitez, A. (2003). La situación demográfica de Venezuela a inicios del tercer milenio. *Temas de Coyuntura, IIES-UCAB, 2003*(47), 45-92.

Gonzalez, G., Gallardo, N., & Bastani, R. (2005). A pilot study to define social support among Spanish-speaking women diagnosed with a breast abnormality suspicious for cancer: A brief research report. *Journal of Psychosocial Oncology*, *23*(1), 109–120. doi:10.1300/J077v23n01_07 PMID:16492647

Granda-Cameron, C. R. N. (1999). The experience of having cancer in Latin America. *Cancer Nursing*, *22*(1), 51–57. doi:10.1097/00002820-199902000-00010 PMID:9990758

Guttman, N. (1997). Ethical dilemmas in health campaigns. *Health Communication*, *9*(2), 155–190. doi:10.1207/s15327027hc0902_3 PMID:11658089

Hanoch, Y. (2004). Improving doctor-patient understanding of probability in communicating cancer-screening test findings. *Journal of Health Communication*, *9*(4), 327–335. doi:10.1080/10810730490468496 PMID:15371085

Hansson, S. O. (2010). Risk: Objective or subjective, facts or values. *Journal of Risk Research*, *13*(2), 231–238. doi:10.1080/13669870903126226

Harvard Medical School's Center of Excellence in Women's Health. (2007). *Black women's attitudes about the links between breast cancer and obesity*. Boston, MA.

Harvey, J. A., & Strahilevitz, M. A. (2009). The power of pink: Cause-related marketing and the impact on breast cancer. *Journal of the American College of Radiology*, *6*(1), 26–32. doi:10.1016/j.jacr.2008.07.010 PMID:19111268

Herbison, S., & Lokanc-Diluzio, W. (2008). The novella approach to inform women living on low income about early breast cancer detection. *Health Promotion Practice*, *9*(3), 294–304. doi:10.1177/1524839906289166 PMID:16923845

Holt, C. L., Lee, C., & Wright, K. (2008). A spiritually based approach to breast cancer awareness: Cognitive response analysis of communication effectiveness. *Health Communication*, *23*(1), 13–22. doi:10.1080/10410230701626919 PMID:18443989

Howze, E. H., Broyden, R. R., & Impara, J. C. (1992). Using informal caregivers to communicate with women about mammography. *Health Communication*, *4*(3), 227–244. doi:10.1207/s15327027hc0403_5

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Hubbell, A. P. (2006). Mexican American women in a rural area and barriers to their ability to enact protective behaviors against breast cancer. *Health Communication*, *20*(1), 35–44. doi:10.1207/s15327027hc2001_4 PMID:16813487

Jandorf, L., Bursac, Z., & Pulley, L. V. (2008). Breast and cervical cancer screening among Latinas attending culturally specific educational programs. *Progress in Community Health and Partnerships: Research, Education, and Action*, *2*(3), 195–204. doi:10.1353/cpr.0.0034 PMID:20208198

Jones, K. O., Denham, B. E., & Springston, J. K. (2006). Effects of mass and interpersonal communication on breast cancer screening: Advancing agenda-setting theory in health contexts. *Journal of Applied Communication Research*, *34*(1), 94–113.

Juraskova, I., Butow, P., Sharpe, L., & Campion, M. (2007). 'What does it mean?' Uncertainty, trust and communication following treatment for pre-cancerous cervical abnormalities. *Psycho-Oncology*, *16*(6), 525–533. doi:10.1002/pon.1104 PMID:16988948

Kahn, J. G., Yang, J. S., & Kahn, J. S. (2010). 'Mobile' health needs and opportunities in developing countries. *Health Affairs*, *29*(2), 252–258. doi:10.1377/hlthaff.2009.0965 PMID:20348069

Kapusta, B., Hryniuk, W., Savage, C., Eaton, G., Yurichuk, S., & Wilcox, B. (2005). Report card 2004. *Cancer Advocacy Coalition of Canada*, *7*(Winter), 1–40.

Kowalski, C., Nitzsche, A., Scheibler, F., Steffen, P., Albert, U., & Pfaff, H. (2009). Breast cancer patients' trust in physicians: The impact of patients' perception of physicians' communication behaviors and hospital organizational climate. *Patient Education and Counseling*, *77*(3), 344–348. doi:10.1016/j.pec.2009.09.003 PMID:19818577

Kreps, G. L. (2008). Strategic use of communication to market cancer prevention and control to vulnerable populations. *Health Marketing Quarterly*, *25*(1), 204–216. doi:10.1080/07359680802126327 PMID:18935885

Kreps, G.L. & Sivaram, R. (2008). Strategic health communication across the continuum of breast cancer care in limited-resource countries. *American Cancer Society, 113*(8), 2331-2337.

Kreuter, M. W., Green, M. C., Cappella, J. N., Slater, M. D., Wise, M. E., Storey, D., & Woolley, S. et al. (2007). Narrative communication in cancer prevention and control: A framework to guide research and application. *Annals of Behavioral Medicine*, *33*(3), 221–235. doi:10.1007/BF02879904 PMID:17600449

Lantz, P. M., Mujahid, M., Schwartz, K., Janz, N. K., Fagerlin, A., Salem, B., & Katz, S. J. et al. (2006). The influence of race, ethnicity, and individual socioeconomic factors on breast cancer stage at diagnosis. *American Journal of Public Health*, *96*(12), 2173–2178. doi:10.2105/AJPH.2005.072132 PMID:17077391

Lee, S. Y., Hwang, H., Hawkins, R., & Pingree, S. (2008). Interplay of negative emotion and health self-efficacy on the use of health information and its outcomes. *Communication Research*, *35*(3), 358–381. doi:10.1177/0093650208315962

Lewens, T. (2010). The risks of progress: Precaution and the case of human enhancement. *Journal of Risk Research*, *13*(2), 207–216. doi:10.1080/13669870903126242

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Lidskog, R. (2008). Scientised citizens and democratized science. Re-assessing the expert-lay divide. *Journal of Risk Research*, *11*(1-2), 69–86. doi:10.1080/13669870701521636

Löfstedt, R. E. (2010). Risk communication guidelines for Europe: A modest proposition. *Journal of Risk Research*, *13*(1), 87–109. doi:10.1080/13669870903126176

Luciani, S., & Andrus, J. K. (2008). A Pan American Health Organization strategy for cervical cancer prevention and control in Latin America and the Caribbean. *Reproductive Health Matters*, *16*(2), 59–66. doi:10.1016/S0968-8080(08)32405-7 PMID:19027623

Luker, K. A., Beaver, K., Leinster, S. J., Owens, R. G., Degner, L. F., & Sloan, J. A. (1995). The information needs of women newly diagnosed with breast cancer. *Journal of Advanced Nursing*, *22*(1), 134–141. doi:10.1046/j.1365-2648.1995.22010134.x PMID:7560521

Macdonald, D. J., Sarna, L., Giger, J. N., Van Servellen, G., Bastani, R., & Weitzel, J. N. (2008). Comparison of Latina and non-Latina white women's beliefs about communicating genetic cancer risk to relatives. *Journal of Health Communication*, *13*(5), 465–479. doi:10.1080/10810730802198920 PMID:18661388

Mairal, G. (2008). Narratives of risk. *Journal of Risk Research*, *11*(1-2), 41–54. doi:10.1080/13669870701521321

McCalman, C. L. (2003). Barriers and motivators for low-income Brazilian women in metropolitan Belo Horizonte: Insights for AIDS prevention. *Health Care for Women International*, *24*(6), 565–585. doi:10.1080/07399330390199410 PMID:12851175

Meyerowitz, B. E., & Chaiken, S. (1987). The effect of message framing on breast self-examination attitudes, intentions, and behavior. *Journal of Personality and Social Psychology*, *52*(3), 500–510. doi:10.1037/0022-3514.52.3.500 PMID:3572721

Mishra, S. I., Chavez, L. R., Magaňa, R. J., Nava, P., Valdez, R. B., & Hubbell, F. A. (1998). Improving breast cancer control among Latinas: Evaluation of a theory-based educational program. *Health Education & Behavior*, *25*(5), 653–670. doi:10.1177/109019819802500511 PMID:9768384

Montenegro, R. A., & Stephens, C. (2006). Indigenous health in Latin America and the Caribbean. *Lancet*, *367*(9525), 1859–1869. doi:10.1016/S0140-6736(06)68808-9 PMID:16753489

Nahon-Serfaty, I. (2005). *Le discours sur la réforme de la santé au Vénézuela: La transition Chávez (1999-2001)*. Unpublished doctoral dissertation, Université de Montréal, Montréal, Canada.

Nápoles-Springer, A. M., Ortíz, C., O'Brien, H., & Díaz-Méndez, M. (2009). Developing a culturally competent peer support intervention for Spanish-speaking Latinas with breast cancer. *Journal of Immigrant and Minority Health*, *11*(4), 268–280. doi:10.1007/s10903-008-9128-4 PMID:18340533

Nekhlyudov, L., & Partridge, A. (2003). Breast cancer risk communication: Challenges and future research directions: Workshop report (United States). *Cancer Causes & Control*, *14*(3), 235–239. doi:10.1023/A:1023684230438 PMID:12814202

OPS. (2011). Género, salud y desarrollo: Venezuela. Retrieved November 11, 2012, from http://www.paho.org/cdmedia/ge_cp/Venezuela.pdf

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Orlando, M.B. & Zúñiga, G. (2000). Situación de la mujer en el mercado laboral en Venezuela: Participación femenina y brecha de ingresos por género. *Temas de Coyuntura, IIES-UCAB*, 41.

Otero-Sabogal, R., Stewart, S., Sabogal, F., Brown, B. A., & Pérez-Stable, E. J. (2003). Access and attitudinal factors related to breast and cervical cancer rescreening: Why are Latinas still underscreened? *Health Education & Behavior*, *30*(3), 337–359. doi:10.1177/1090198103030003008 PMID:19731500

Ponce, M. G. (2009). *La pobreza en Venezuela: Mediciones y diversidades. Presentación realizada en el taller sobre salud, mujer y telefonía celular*. Caracas, Venezuela: UCAB.

Quillin, J. M., McClish, D. K., Jones, R. M., Wilson, D. B., Tracy, K. A., Bowen, D., & Bodurtha, J. N. et al. (2008). Duration of an intervention's impact on perceived breast cancer risk. *Health Education & Behavior*, *35*(6), 855–865. doi:10.1177/1090198108325912 PMID:19011219

Robles, S. C., & Galanis, E. (2002). Breast cancer in Latin America and the Caribbean. *Pan American Journal of Public Health*, *11*(3), 178–185. doi:10.1590/S1020-49892002000300007 PMID:11998184

Roeser, S. (2010). Intuitions, emotions and gut reactions in decisions about risks: Towards a different interpretation of 'neuroethics'. *Journal of Risk Research*, *13*(2), 175–190. doi:10.1080/13669870903126275

Rogers, C., Thompson, K., & Robinson, S. (2002). Introducing a breast health strategy into schools. *Health Education*, *102*(3), 106–112. doi:10.1108/09654280210425994

Salant, T., & Gehlert, S. (2008). Collective memory, candidacy, and victimisation: Community epidemiologies of breast cancer risk. *Sociology of Health & Illness*, *30*(4), 599–615. doi:10.1111/j.1467-9566.2007.01079.x PMID:18298621

Schutt, R. K., Cruz, E. R., & Woodford, M. L. (2008). Client satisfaction in a breast and cervical cancer early detection program: The influence of ethnicity and language, health, resources, and barriers. *Women & Health*, *48*(3), 283–302. doi:10.1080/03630240802463475 PMID:19064463

Schwartsmann, G. (2001). Breast Cancer in South America: Challenges to improve early detection and medical management of a public health problem. *Journal of Clinical Oncology*, *19*(18), 118s–124s. PMID:11560986

Shaw, B. R., McTavish, F., Hawkins, R., Gustafson, D. H., & Pingree, S.Bret R. Shaw, Fiona McTavish, Rober. (2000). Experiences of women with breast cancer: Exchanging social support over the CHESS computer network. *Journal of Health Communication*, *5*(2), 135–159. doi:10.1080/108107300406866 PMID:11010346

Small, S. A. (1995). Action-oriented research: Models and methods. *Journal of Marriage and the Family*, *57*(4), 941–955. doi:10.2307/353414

Small, S. A., & Uttal, L. (2005). Action-oriented research: Strategies for engaged scholarship. *Journal of Marriage and the Family*, *67*(4), 936–948. doi:10.1111/j.1741-3737.2005.00185.x

Smith, S. W., Nazione, S., Laplante, C., Kotowski, M. R., Atkin, C., Skubisz, C. M., & Stohl, C. (2009). Topics and sources of memorable breast cancer messages and their impact on prevention and detection behaviors. *Journal of Health Communication*, *14*(3), 293–307. doi:10.1080/10810730902805903 PMID:19440911

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Stoffle, R., & Minnis, J. (2008). Resilience at risk: Epistemological and social construction barriers to risk communication. *Journal of Risk Research*, *11*(1-2), 55–68. doi:10.1080/13669870701521479

Street, R. L. (2003). Communicating in medical encounters: An ecological perspective. In T. L. Thompson, A. M. Dorsey, K. I. Miller, & R. Parrot (Eds.), *Handbook of health communication* (pp. 63–89). London: Lawrence Erlbaum Associates.

Strydom, P. (2008). Risk communication: World creation through collective learning under complex contingent conditions. *Journal of Risk Research*, *11*(1-2), 5–22. doi:10.1080/13669870701521248

Visschers, V. H. M., Meertens, R. M., Passchier, W. F., & de Vries, N. K. (2008). Audiovisual risk communication unraveled: Effects on gut feelings and cognitive processes. *Journal of Risk Research*, *1*(1-2), 207–221. doi:10.1080/13669870801947954

Wang, J. H., Liang, W., Schwartz, M. D., Lee, M. M., Kreling, B., & Mandelblatt, J. S. (2008). Development and evaluation of a culturally tailored educational video: Changing breast cancer-related behaviors in Chinese women. *Health Education & Behavior*, *35*(6), 806–820. doi:10.1177/1090198106296768 PMID:17602099

Williams, T., Clarke, V. A., & Savage, S. (2002). Women's perceptions of familial aspects of breast cancer. *Health Education*, *102*(2), 50–59. doi:10.1108/09654280210418983

Winzelberg, A. J., Classen, C., Alpers, G. W., Roberts, H., Koopman, C., Adams, R. E., & Taylor, B. C. et al. (2003). Evaluation of an Internet support group for women with primary breast cancer. *Cancer*, *97*(5), 1164–1173. doi:10.1002/cncr.11174 PMID:12599221

Wonderlick, A. L., & Fine, B. A. (1997). Knowledge of breast cancer genetics among breast cancer patients and first-degree relatives of affected individuals. *Journal of Genetic Counseling*, *6*(2), 111–130. doi:10.1023/A:1025651816768

Yamin, A. E. (2000). Protecting and promoting the right to health in Latin America: Selected experiences from the field. *Health and Human Rights*, *5*(1), 116–148. doi:10.2307/4065225 PMID:11154526

ENDNOTES

1 These articles indicate the necessity of state intervention to improve healthcare: Article 83 describes health as a fundamental social right that is the responsibility of the state; Article 84 describes health as a guarantee for all through the state's creation and management of a national health system; and Article 85 highlights that the state will regulate and finance the public health system (Feo & Siqueira, 2004).

2 This action-oriented research project was supported by research grants from the Canadian Institutes of Health Research. It also received additional funding from private-sector donors in Venezuela such as Avon Venezuela, Roche, and Novartis.

3 This four-year (2009-2012) action-oriented research project was developed by Professor Isaac Nahon-Serfaty and Professor Mahmoud Eid from the University of Ottawa (Canada) in partnership with the Venezuelan Breast Cancer NGO, SenosAyuda, with support from Professors Andrés Ca-

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

ñizales, Alfredo Poggi, José Rafael Briceño, Acianela Montes de Oca, and Néstor Garrido from the School of Social Communication of the Universidad Católica Andrés Bello (Caracas, Venezuela), and the Colegio Nacional de Periodistas, the association representing professional journalists in Venezuela.

[4] A total of 121 people participated in the four workshops (June 5-6, 2009; June 11-12, 2010; May 20-21, 2011; and February 24-25, 2012). They were 50 activists representing different NGOs and including survivors and patients, 23 physicians, 28 journalists, 9 scholars, 5 public officials, and 6 representatives from the private sector.

This work was previously published in the International Journal of Risk and Contingency Management (IJRCM), 4(3); edited by Kenneth David Strang, pages 49-69, copyright year 2015 by IGI Publishing (an imprint of IGI Global).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 24
The Role of Primary Health Care in Prevention, Early Detection, and Control of Cancer

Salwa Bahkali
King Saud Bin Abdulaziz University for Health Sciences, Saudi Arabia

Amin Bawazir
King Saud Bin Abdulaziz University for Health Sciences, Saudi Arabia

Mowafa Househ
King Saud Bin Abdulaziz University for Health Sciences, Saudi Arabia

Mohamed Khalifa
King Faisal Specialist Hospital and Research Center, Saudi Arabia

Mohamud Sheikh
University of New South Wales, Australia

ABSTRACT

Cancer is one of the leading causes of death worldwide, and the incidence is growing. Recent evidence shows a reduced risk of dying from cancer. For years Primary Health Care (PHC) has played a vital role in promoting health, but little has been done in emphasizing its role in reducing the incidence of and mortality from cancer through performing early diagnosis. PHC is directly involved in the initial diagnosis of more than 85% of all cancer cases worldwide (Vedsted & Olesen, 2009). PHC also has an important role in the public awareness about the importance of screening, especially in high-risk patient groups. The interaction between the patient and the health service is crucial in ensuring that relevant alarming symptoms are presented and that action is taken at the earliest possible time. This chapter aims to explore the role of primary healthcare in the prevention, early detection, and control of cancer in a developing nation - Saudi Arabia.

INTRODUCTION

The global burden of non-communicable diseases (NCDs) is constantly increasing. According to the World Health Organization (WHO), the spread of these diseases presents a global crisis in almost all countries (WHO/PHAC, 2005). The increasing global crisis in NCDs is a barrier to the Millennium Development Goals (MDG) including poverty reduction, health equity, economic stability, and human

DOI: 10.4018/978-1-5225-0549-5.ch024

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

security (Beaglehole et al., 2011). In addition, NCDs are considered a major cause of health inequality (WHO, 2008). The underlying causes of these diseases share modifiable risk factors that are well known and are similar in many countries (Yusuf et al., 2004). In recent years, substantial gains have been achieved all over the world in economic growth, health, and living standards in the past century. This progress is now threatened by the substantial increase of NCDs, especially heart diseases, cancers, strokes, diabetes, and chronic respiratory diseases (Engelgau, 2011).

The United Nation General Assembly (UNGA) held a special session in 2011 to tackle NCDs' burden and mobilize the political leadership to optimize their coordination and consensus for priority actions and interventions in responding to the crisis of NCDs (Hospedales et al., 2011). It was also directed to enhance the development of strategic decisions to reorient population-based prevention, in addition to clinic and hospital-based care policies toward NCDs including cancers (Engelgau et al., 2010). How much focus should be given to the prevention of chronic diseases and cancer, and how much focus should be given to the treatment of those already affected by such diseases is an important policy question that is usually directed to government decision makers everywhere around the world (WBO, 2012).

As cancer is one of the major NCDs, and it is well known to be one of the world's top reasons for mortality, the World Health Organization (WHO) and other organizations are concerned about the integration of cancer prevention, early detection, treatment and palliative care within the context of the National Non-communicable Diseases prevention and control activities at the level of the primary healthcare settings (WHO, 2008). Furthermore, cancer is one of the major cause of premature death worldwide. For example, according to the World Cancer Report in 2014, 14 million new cases and 8 million death cases have been reported (Steward & Wild, 2014). The new number of cases is expected to reach 15 million by 2020 (WHO, 2003), where the developing countries account for 53% of the new cancer cases and 60% of cancer-related deaths. The economic value of disability-adjusted life-years (DALYs) - loss due to cancer amounted to 895 billion USD in 2008 globally. This represents approximately 1.5% of the world's GDP (Busse et al., 2010).

In Middle Eastern countries 467,000 new cases and 323,000 deaths from cancer were reported in 2008 (Boyle & Levin, 2008). The rate increases as communities adopt western lifestyles and become urbanized. In addition, the steady increase in the population's expected age is attributed to be one of the key factors that contribute to the growing occurrence of diseases. Some factors are considered in contributing to the ongoing increment of cancer which include, among others, population ageing; globalization of risks; poverty, which increases acutely in less developing countries; in addition to the longstanding challenges of infectious diseases (Steward & Wild, 2014). These factors double the burden of diseases, which place enormous strains on resource poor health systems.

Remarkable and sustainable National Cancer Control Programs (NCCPs) were developed and implemented in the Kingdom of Saudi Arabia(KSA) and other Gulf Cooperation Council (GCC) countries. Their priority goal is to enhance the public health approaches for the prevention and control of cancer with a special focus on primary healthcare, which fits into the broader WHO framework to support health systems and act as part of the action plan implementation for the Global Strategy of Prevention and Control of NCDs, which was endorsed by the World Health Assembly in May 2008 (Al-Eid, 2007) and emphasized in 2011 (WHO, 2008). The primary healthcare system in many countries including Saudi Arabia is still based on providing management of acute cases and follow-up care rather than initiating prevention and early detection of chronic NCDs including cancer (Pruitt & Epping-Jordan, 2005). A patient referral system is not in place and government policy allows self-referral on demand to secondary

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

and tertiary facilities. Thus, patients often use these higher-level facilities because they are aware that primary care facilities lack the capacity to manage NCDs in terms of their ability to perform clinical investigations and provide proper medications.

THE GLOBAL BURDEN OF CANCER

Cancer is one of the main causes of death everywhere in the world. In 2011, around 7.9 million people died from cancer worldwide, which is approximately 13% of the total global deaths. However, it is presumed that the number of deaths from cancer will increase to reach 12 million by 2030 (WHO, 2013). Worldwide, breast, colorectal, and lung cancer are the most common and leading causes of cancer death among women, whereas lung, prostate, colorectal, and liver cancer are the leading cause of cancer death among men (NCI, DHHS & CDC, 2013). Liver cancer is the third most common cause of cancer mortality, with nearly 550,000 annual deaths (Joseph et al., 2006). The American Cancer Society estimates 14,270 deaths from liver cancer in the US in 2004, of which 70%-80% were hepatocellular carcinoma (HCC). Liver cancer is the eighth most frequent cause of cancer mortality in US males and the twelfth in females (Evans, London, & Engstrom, 2005).

In the South-West Asia region, the most widespread cancers in males vary, with lung, urinary, bladder or liver in first place, while for females breast cancer is the most diagnosed NCD. In both genders, non-Hodgkin's lymphomas and leukemia are relatively prevalent, along with lung in males and thyroid in specific female populations (Salim et al., 2010). The Gulf Cooperation Council (GCC) estimates that the overall occurrence of cancer in the Gulf region between 1997 and 2007 was 95,183 cases. Almost half of the cases (47%) were presented with advanced tumors (either regional or distant metastasis), only 22% presented with localized tumor, and 1% was in situ, while 30% of the cases were presented with unknown extent of cancer (Al-Madouj, Eldali & Al-Zahrani, 2011).

The Role of Preventable Communicable Diseases in the Incidence of Cancer

Evidence reveals that there is a link between some of the preventable communicable diseases such as the hepatitis C virus (HCV), hepatitis B virus (HBV), and human papilloma virus (HPV) with cancer disease. According to WHO, Hepatitis B is a potentially life threatening liver infection that can cause chronic liver disease and cause death as a result of liver cancer and cirrhosis (WHO, 2012).

Hepatitis B and C viruses are known to be the major risk factors for hepatocellular carcinoma (HCC) worldwide. Chronic infection with HCV and/or HBV is associated with the majority of HCC cases. Recent estimates of the incidence in the U.S. have attributed 47% of cases to HCV alone, 15% to HBV alone, and 5% to co-infection with both viruses (Wong & Goh, 2006; Henry, Bosch & Bowers, 2002; El-Serag 2007; Parkin, 2011; Ayoola & Gadour, 2004). Furthermore, some epidemiological studies suggest a link between hepatitis C virus (HCV) infection, which infects more than 200 million people worldwide, and some B-cell non-Hodgkin's lymphomas. Treatment with interferon in these patients can lead to regression of the lymphoma (Hermine et al., 2002; Mele et al., 2003; Mazzaro, Tirelli, & Pozzato, 2005). Moreover, evidence reports a significant reduction of HCC as a result of reductions in the rate of chronic HBV infection due to the newborn and infant regular vaccination (Tennant et al., 2004).

Many studies describe the carcinogenic effect of human papillomavirus (HPV) on the head and neck squamous cell cancers (HNSCC) particularly in tonsillar cancer (Mellin et al., 2000; Dayyani et

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

al., 2010; Marur et al., 2010) and on cervical intraepithelial neoplasia (CIN). HPV-16/18 infection is projected to account for 70% of all cervical cancers worldwide, but is slightly higher in more developed (72–77%) than in less developed (65–72%) regions (Clifford, et al., 2006; Altaf, 2006). It was also found that HPV vaccine reduces the risk of HPV related cancer, particularly CIN. Besides, there is a visible cross-protection effect (20%) to the other non-vaccine type of HPV on cervical cancer (Brown et al., 2009). However, vaccination is most effective when applied to young women before they become sexually active (Sait, et al., 2012).

The Burden of Cancer in the Kingdom of Saudi Arabia

According to published cancer statistics, the Saudi Cancer Registry (SCR) reported 12,309 new cancer cases diagnosed in Saudi Arabia in 2007. King Faisal Specialist Hospital and Research Centre (KFSHRC) reported 2,527 new cancer cases during the year 2011which constitute about 23% of the total cancer cases reported to SCR by all the hospitals in the Kingdom (CTR et al, 2012). However, this figure is expected to increase to 30,000 cases within 15 years.

The age-adjusted occurrences of cancer in Saudi Arabia were in a low rate compared to the remaining Arab countries. It accounts for 60 females and 61 males per 100,000. Although incidences are lower than in the West, breast cancer occupied the first place in Saudi Arabia as other countries. Colorectal cancer (CRC) is the most common cancer in Saudi males and relatively common in females. Patients are diagnosed with CRC at more advanced stage of the disease and at a younger age (30-74 years) compared to Western populations (Al-Jebreen, 2007; Ibrahim et al., 2008). Breast cancer is the most common cancer in women. Both ranked in the first position for all age group (Al-Eid, 2007; El-Attar, 2005; Alam, 2006; Dey, 2010).

Thyroid cancer is regarded as the most frequent endocrine malignancy in Saudi females, with a variable geographic and ethnic occurrence around the country. It comes in the second position in the list for the female group (30-59 years), and the third for the male group (30-44years) (Al-Eid, 2007; Qari, 2004).

Non-Hodgkin's lymphoma and leukemia are more common in the pediatric and young adult group (Al-Eid, 2007), while lung cancer is the most common male cancer after the age of 45 years; whereas, it become more common with advanced age (Al-Eid, 2007; Salim et al., 2010)

Liver cancer is the most common cancer in elderly males. It is ranked as number one after the age of 75 years and it moves down in the list as age decreases (Al-Eid, 2007). The occurrence of genito-urinary cancer (GUC) in Saudi Arabia is comparatively low. It accounts for 9.2% of all cancers in Saudi Arabia. It increases with age; however, males were 5-fold higher than females for the occurrence of genito-urinary cancer. The most common GUC in males is bladder, prostate, kidney, and testicular cancer respectively (Abomelha, 2004; Taha & Kamal, 2005 ; Dandash & Al-Mohaimeed, 2007).

The cervical cancer rate is lower than expected in Saudi Arabia. This may refer to underreporting of the cases. Current estimates in Saudi Arabia indicate that every year, 152 women are diagnosed with cervical cancer and 55 die from the disease. It is predicted that as the population ages, there will be a dramatic increase in the incidence of cervical cancer. The estimated number of new cervical cancer cases and deaths in 2025 is 309 and 117, respectively. Although cervical cancer is both preventable and curable, most women in Saudi Arabia present at advanced stages due to the lack of effective screening program (Sait, et al., 2012). Skin cancer is a common malignant neoplasm in Saudi Arabia after the age of 60, where most patients are diagnosed with mostly basal cell carcinoma (BCC) and squamous cell carcinoma (SCC) followed by Kaposi's sarcoma. In contrast to the western countries, melanoma

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

is a rare skin neoplasm in Saudi Arabia (Al-Eid, 2007; Elsayed et al., 2010; Maghrabi, AL-Ghamdi, & Elhakeem, 2004). Oral cancer is relatively rare in Saudi Arabia; however, there are very wide regional disparities in occurrence of oral cancer among the different regions. It is a significant public health problem for the residents of Jizan and the women of Najran (Salim et al., 2010; Brown, Ravichandran & Warnakulasuriya, 2006).

Most of the cancer prevention screening programs are not in place or not effective. (Salim et al, 2009). The preventive measures for breast cancer, which is the most frequent cancer in women, are probably less than expected. There is an imbalance between the knowledge and practice at all levels. Hence, greater focus on breast cancer education at community-based programs is needed to improve the knowledge, and change patient attitude and behavior rectify the misconceptions towards cancer and its risk factors (Alam, 2006; Dandash & Al-Mohaimeed, 2007). Several studies reported that higher survival rates of preventable cancer, such as breast or cervical cancer, could be attainable through: increased awareness of cancer and of the prospective for successful treatment; a high-quality primary care system without economic or cultural barriers to access; and a well-functioning referral system for basic surgical, medical, or radiotherapy treatments (Sait et al., 2012; Shulman et al., 2010).

Economic Burden of Cancer

Analysis of worldwide cancer registries reveals that males in Eastern Europe had the largest cancer burden (3,146 age-adjusted DALYs lost per 100,000 men), while sub- Saharan Africa females had the highest burden (2,749 age-adjusted DALYs lost per 100,000 women) in the world (US health news, 2012). Breast, colorectal, lung and prostate cancer were the notable contributors to total DALYs in most areas, accounting for 18%-50% of total cancer burden. Infection-related cancers such as liver, stomach and cervical cancer accounted for a larger part of overall DALYs in eastern Asia (27% of all cancers) and in sub-Saharan Africa (25% of all cancers) than in other regions.

Saudi Arabia is one of the world's 25 largest economies (24th) and No. 1 in the MENA region (RCJY, 2011). One principal area that the Saudi authorities focus attention on is the provision of healthcare services to its citizens. Hence, the amount of government expenditures allocated for healthcare services is constantly increasing. According to the World Bank, healthcare expenditures in the Saudi Arabia represented 4.3 percent of the GDP in 2010. The share of public health contributed approximately 2.7 to the GDP, while the private sector represented 1.6 percent. Healthcare expenditure in Saudi Arabia exceeds SR 91.20 billion (USD 24.35 billion), which means an increase of 16% during 2011. The expenditure on the health sector will reach, according to NCB Capital's predictions, to SR174 billion for 2017 (Gazette, 2012). GCC countries including Saudi Arabia will face an extraordinary rise in demand for healthcare over the next two decades. It is predicted that the total healthcare spending in the region will reach US$60 billion by 2025 (Mourshed, Hediger & lambert, 2007).

Despite the low cancer incidence reports in Saudi Arabia, the country must be ready to face the challenges of predictable increases in the cancer burden which is attributed to growth and aging of its population. Therefore, future cancer rates might demonstrate a significant increase and place large demands on healthcare resources (Ibrahim et al., 2008). The KSA is taking the initiative for addressing this expected increase in cancer incidence by establishing multiple contemporary facilities in many regions where treatment of all levels of care can be given free of charge with the highest international quality of care standards. Many cancer centers have been established in major and smaller cities in KSA, with several more underway. The Kingdom has made a significant investment in training physicians in oncology.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Furthermore, there are multiple governmental and nongovernmental service for cancer control such as screening and early-detection programs for certain cancers.

However, cancer in Saudi Arabia is generally diagnosed at an advanced stage, which increases the burden on the health system. The fact is that the shortage of health care resources and constraints make it difficult for most countries to provide comprehensive cancer prevention and control services particularly in the public health sector. Most of the healthcare system funds for cancers are in tertiary care without a balanced investment in prevention, early detection, and palliative care (Hixon & Maskarinec, 2008). The increase in pharmaceuticals spending was estimated to be 13.5% (from SR16.70 billion (USD 4.46 billion) in 2011 to SR18.95 billion (USD 5.06 billion) in 2012). The expenditure on medical equipment was estimated to have increased from SR5.54 billon (USD 1.48 billion) in 2011 to SR6.53 billion (USD 1.74 billion) in 2012 (Howard, 2014).

ROLE OF PRIMARY HEALTHCARE IN CANCER PREVENTION AND CONTROL

The Alma-Ata Declaration of 1978 emerged as a major milestone of the twentieth century in the field of public health and it considered primary healthcare as the key to achieve the goal of Health for All (Anderson et al., 2006). The declaration has defined the role of the primary healthcare functioning as "an essential healthcare based on practical, scientifically sound and socially acceptable methods and technology which has become universally accessible to individuals and families in the community through their full participation and at a cost that the community and the country can afford to maintain at every stage of their development in the spirit of self-reliance and self-determination" (MOH, 2002).

The role PHC centers play in health promotion to provide health education is duly recognized. The concept of PHC is to provide the basic health services for all members of the community, and represents the first level of the community's contact with the health services (Al-Mazrou, 2002). PHC providers believe that people are the protectors of their own health and at the same time they are the resources. Users and decision makers regarding the use of health information take decisions about the activities to be undertaken to ensure fulfillment of health (Almalki, Fitzgerald & Clark, 2012).

Since the Alma-Ata declaration, the Saudi Ministry of Health (MOH) decided to activately pursue the development of the preventive health services by adopting the PHC approach as one of its key health strategies. Consequently, in 1980, a ministerial decree was issued to establish the PHC centers as part of national health services aimed at helping people to live healthy, safe and independent lives (Midhet, Al Mohaimeed & Sharaf, 2010; Samb, Bolhuis & Etienne, 2011).

Evidence shows that the future of PHC depends on the supported ongoing evaluation of all established and implemented programs at PHC centers; the extensive use of health education approaches and resources; the needs, demands, expectations, and acceptance of the Saudi community; the implementation of the 'health insurance' policy; qualifying and training of the primary healthcare providers to enable them to understand and cope with current methods of healthcare provision (Al-Mazrou, 2002).

Ensuring early diagnosis and palliative care through a PHC approach is one of the most feasible strategies to increase early detection, improve survival rates, and respond to the urgent needs of patients at advanced stages of the disease (Hixon & Maskarinec 2008). Family physicians play an important role in the health promotion and cancer screening programs, as well as the dynamic actions in following-up cancer patients vigorously after their return from oncology facilities. Eliminating NCD risk factors such

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

as unhealthy lifestyle, enhancing tobacco control, and managing obesity are the most important measures for effective control of most cancers in PHC setting (Ibrahim et al., 2008; Dey, 2010).

The improvement action of PHC on NCDs in general, cancer in particular, focuses on the delivery of better quality primary care services, places greater emphasis on addressing the constraints and barriers in healthcare services to optimize the work and to achieve universal coverage. Good quality primary care services are up-to-date with the core values and principles of primary healthcare, offer great potential for improving NCDs prevention and control, along with meeting the need for integrated management of NCDs, irrespective of the etiology (WHO, 2012).

Disparities in health and healthcare will limit the continuous improvement of the overall quality of the care provided which will negatively affect the health of the population negatively, and ultimately result unnecessary costs. A "healthcare disparity" refers to differences in access to, quality, or availability of healthcare facilities and services between socioeconomic and/or geographically defined population groups (Mourshed, Hediger, & Lambert, 2007; RCJY, 2011).

In 2012, Aldakheel reported that PHC services were most effective in the prevention program of communicable diseases such as immunization and vaccination. However, PHC services were less effective when dealing with non-communicable diseases and preventive programs. Besides, there was absenteeism of evidence-based practice and disorganization for most of the efforts provided by healthcare services (Aldakheel, 2012).

AL-Ahmadi reviewed 31 published literatures in 2006 on the quality of primary health care services in Saudi Arabia and gave poor scores to the access and effectiveness of chronic disease management programs, health education, and referral systems. Several factors were identified as barriers including: management and organizational factors; poor implementation of evidence-based practice; inadequate professional development; misuse of referrals to secondary care; and different organizational culture (AL-Ahmadi & Roland, 2006). Furthermore, healthcare providers (HCPs) are one of the main human resources in PHC centers. Shortage of qualified HCPs will affect the quantity and the quality of service delivered in PHC centers to prevent, early detection, or control NCDs including cancer. Satisfaction in the work environment is one of the determinants for preservation of skilled physicians and nurses. Multiple studies were conducted in PHC centers in different regions in Saudi Arabia, which reported that the majority of the HCPs were not satisfied with their jobs especially in rural areas. Accordingly, being unsatisfied caused a negative impact either directly or indirectly on clinical and non-clinical staff motivation, productivity, and performance. The major influencing factors for poor productivity were unsuitable working hours/shifts, lack of facilities, inability to match work with family needs, inadequacy of family-leave time, poor staffing, management and supervision practices, lack of professional development opportunities, and inappropriate working environment in terms of the level of security, patient care supplies and equipment (ALJuhani & Kishk, 2006; Almalki, 2012).

Several studies were conducted at the majority of the PHC centers in the Asir region - one of Saudi Arabia's rural areas, which revealed that there was a shortage of many essential resources for NCDs prevention and management in that region (Al-Sharif & Al-Khaldi et al., 2002). Fewer health professionals have skills required to manage PHC services, using evidence-based medicine (EBM). Al-Baghlie & AI-Aimaie assessed the attitudes of 409 physicians towards the use of EBM approaches in the Dammam area of eastern Saudi Arabia and have reported that only 108 physicians (39.6%) had heard about EBM. However, the overall attitude toward EBM approaches was positive (AI-Baghlie & AI-Aimaie, 2004).

Having an effective referral system is an important strategy for the delivery of high quality, comprehensive, and integrated healthcare services. Various studies have been conducted in different Saudi

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

regions to assess the referral system. The studies show poor quality of the referral system process although some referral processes were of good quality. The following factors contribute to poor patient referral: incomplete forms, unclear handwriting, poor medical skills, carelessness, insufficient space provided for history, lack of communication, and lack of feedback notes to the PHC centers (Baghdadi L & Baghdadi R, 2007).

Primary care is the backbone of disease prevention. It is an integral part of the National Healthcare System, of which it is the core, as well as of the global, social, and economic development of the community. It represents the first contact of individuals, families and the community with the National Healthcare System, bringing medical care as close as possible to the place where people live and work, and it is the first link in the entire process of medical attention.

Unger et al, have evaluated the impact of each disease control program on local healthcare facilities and have identified how the integration of these programs affects the quality of healthcare delivery and its utilization of communities. They found that in order to take advantage of appropriate integration and prevent disease control programs from damaging general care delivery, disease control activities should generally be integrated with health centers which offer patient-centered care, non-profit health facilities, and avoid conflict with healthcare delivery. The authors recommend the development and implementation of best practices used for governmental and international aid organizations (Unger, Paepe & Green, 2003).

Unfortunately, screening programs in Saudi Arabia are limited but on the rise. Although, breast cancer is the most frequent cancer in diagnosed in Saudi women, the preventive measures for such problem are probably less than expected. Even if the knowledge on breast cancer preventive measures including screening is improved, there is a mismatch between the knowledge and practice of self-breast examination (SBE) among women. Hence, greater focused breast cancer education that is community-based program is needed to improve the knowledge and change the patient's attitude and behavior and the mistaken belief towards cancer and its risk factors (Alam, 2006; Dandash & Al-Mohaimeed, 2007)

Higher survival rates of breast cancer can be attained through increased awareness of breast cancer and of the prospective for successful treatment, a high-quality primary care system without economic or cultural barriers to access, and a well-functioning referral system for basic surgical and hormonal treatment are mainly needed (Farmer et al., 2010). Pap smear screening has helped to reduce cervical cancer occurrence and mortality rates by 70% in developed countries (Sait et al., 2012).

The cancer burden in Saudi Arabia is likely to increase more as the population ages. The current health system capacity needs significant support to meet the growing cancer burden. Due to the lack of studies in Saudi Arabia regarding the role of PHC in prevention, early detection and control of cancers, there is an urgent need to assess the current capacity of PHC centers with respect to cancer control and prevention to know how Saudi Arabia is progressing and to open the doors for other researchers and policy makers to step forward to develop and empower the PHC centers to accommodate the huge number of cases.

FUTURE DIRECTIONS

The World Health Organization began promoting PHC since 1978, when PHC was adopted during the Alma-Ata Conference as a central strategy to achieve the "Health for All" goal. Recently, a process to renew the PHC initiative was initiated by Pan American Health Organization/World Health Organization (PAHO/WHO) with the aim to strengthen countries' potential to implement a coordinated, effective and

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

sustainable strategy to tackle the existing health problems, meet new health challenges, and improve health equity (PAHO & WHO, 2012).

The emergence of NCDs as a major health problem may need fundamental reorientation in the way PHC is delivered in the Saudi healthcare system. Moreover, the expansion and improvement of the preventive services of PHC requires providing better quality care where PHCs must become part of an integral process to respond to cancer control and prevention.

The national and regional goals of prevention, early detection, and control of cancer should be in line with the international principles shared with Health 2020 that will guide all actions in the cancer prevention, early detection strategic plan, from priority-setting through implementation to evaluation (Adeyi, Smith & Robles, 2007). Early detection of cancer is the key to survival; therefore, cancer should be detected during periodic health examinations, either through screening procedures or through the observation of signs and symptoms, which prompt a routine visit to a primary care physician.

The ongoing assessment of primary healthcare service is one of the vital issues for improved health promotion, cancer prevention, early detection and integrated care. It should be done to assure: equity, improvement and the strengthening and development of primary healthcare services; empowerment of all PHC centers' activities to enhance, but not replace community action, and to promote health literacy; integration of PHC centers with non-health sector to provide a health-supporting environment that promotes coping with disability, social protection and appropriate and accessible social and healthcare services; evaluate and improve the integrated programs that attain a multiple intervention strategy for all NCDs rather than individual interventions so as to assure cost effectiveness.

Vedel et. al, has summarized the barriers of cancer screening related to the physician, the patient and the healthcare system where the primary care physicians do not believe in the usefulness of cancer screening and the patients feel embarrassment, discomfort or fear of the test (Vedel et al, 2011).

To assess whether or not the PHC organization is ready for such activity, it is necessary to define and measure its potentials regarding cancer prevention, early detection, and control. This is in addition to conducting an assessment of its ability of care continuity with the existence of the conceptual and practical difficulties that they face. Information obtained from this assessment will provide a better understanding of the main barriers facing the process of implementing cancer prevention, early detection and control strategies in Saudi Arabia's primary healthcare facilities.

Data obtained from the assessment of the primary healthcare facility (Facility-based information) will inform policy makers and stakeholders with the actual information on the reality of the level of service delivery (input, process, costs, output and quality). Such information is needed for monitoring, evaluation and improving the quality and performance at the facility-level. Moreover, the performance and quality factors can affect patients' health-seeking behavior (e.g., utilization of services) and can also mediate the impact of service utilization on population-level outcome measures.

CONCLUSION

Developing nations such as Saudi Arabia require qualified and well trained manpower in addition to effective operating systems, patient education and effective disease prevention. This also needs well oriented community primary health care, evidence-based medicine, patient access to secondary and tertiary care, professional practice and on-going professional education, in addition to proper training,

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

better services utilization review and safety inspections. All this will incorporate into prevention, early detection and control of health problems under the auspices of PHCs.

A disproportionate focus on specialist, tertiary care encourages policy-makers, funders and service users alike to by-pass primary care provision in favor of hospital-based treatment. Fragmentation resulting from duplication of programs and projects is shaping health systems in a way that weakens service delivery models that are efficient and equitable. The growing commercialization of healthcare in unregulated health systems makes integration of services difficult and closes the door of healthcare for the poor. Government policy and resources are important pillars working to ensure quality of care across all PHC sectors and to integrate the fragmented provision of care for NCDs through strong health sector planning and appropriate regulation. However, effective delivery of good quality care in primary level settings depends on a strong health workforce that is appropriately educated and trained and equipped with a good supervision and dependable structures for referral to secondary and specialist tertiary care as needed. However, primary care services will not be able to meet the needs of people with NCDs, especially cancers, without reliable and affordable access to essential medicines, and other technologies of good quality, over an extended period of time (Farmer, 2010); medicines are an essential component of treatment of many cancers and/or palliative treatment (Santosa & R.N., 2008).

It is of a paramount importance to establish resources, policies and procedures on developing universal good quality, patient centered primary healthcare with referral to secondary and tertiary levels of care whenever needed. It is essential to establish future strategic plans for the function of primary healthcare for prevention, early detection and control of cancer. The role of primary health care in screening high risk patient groups is very crucial to the process of prevention of the disease and early detection of cancer cases. Healthcare policy makers should be focusing on investing in primary healthcare services as a first point of contact with the public for both awareness and early detection of cancer. We need to emphasize and support the role of primary healthcare in reducing the incidence of, and mortality from, cancer by early diagnosis through developing and implementing cancer screening programs and through making these programs accessible for the public and increasing the public awareness about their existence and importance.

REFERENCES

Abomelha, M. (2004). Genito-urinary cancer in Saudi Arabia. *Saudi Medical Journal*, *25*(5), 552–556. PMID:15138519

Adeyi, O., Smith, O., & Robles, S. (2008). Public Policy and the Challenge of Chronic Non-communicable Diseases. *International Journal of Epidemiology*, *37*(3), 686–687. doi:10.1093/ije/dym222

AI-Baghlie, N. & AI-Aimaie, S. (2004). Physician attitudes towards evidence-based medicine in eastern Saudi Arabia. *Annals of Saudi Medicine*, *24*(6). PMID:15646158

AL-Ahmadi, H. & Roland, M. (2006). Quality of primary healthcare in Saudi Arabia. *International Journal for Quality in Health Care*, *17*(4), 331–346.

Al-Eid, H. (2010). *Cancer incidence and survival report Saudi Arabia 2007*. Kingdom of Saudi Arabia Ministry of health Saudi cancer registry.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Al-Jebreen, A. (2007). Clinico-pathological patterns of colorectal cancer in Saudi Arabia: Younger with an advanced stage presentation. *Saudi Journal of Gastroenterology*, *13*(2), 84–87. doi:10.4103/1319-3767.32183 PMID:19858619

AL-Juhani, A. & Kishk, N. (2006). Job satisfaction among primary healthcare physician and nurses in AL-Madinah AL-Munawwarah. *The Journal of the Egyptian Public Health Association*, *81*(3-4), 165–180. PMID:17382059

Al-Khaldi, Y., Al-Sharif, A., Al-Jamal, M., & Kisha, A. (2002). Difficulties faced when conducting primary healthcare programs in rural areas. *Saudi Medical Journal*, *23*(4), 384–387. PMID:11953760

Al-Madouj, A., Eldali, A., & Al-Zahrani, A. (2011). *Ten-year cancer incidence among nationals of the GCC states 1998-2007*. Gulf center for cancer control and prevention. Retrieved from: http://www.moh.gov.bh/pdf/publications/GCC%20Cancer%20Incidence%202011.pdf

Al-Mazrou, Y. (2002). Primary healthcare in Saudi Arabia: Its development and future prospective. *J Family Community Med*, *9*(2), 15–16. PMID:23008666

Al-Sharif, A. & Al-Khaldi, Y. (2003). Resource availability for care of hypertensive at primary health settings in Southwestern Saudi Arabia. *Saudi Medical Journal*, *24*(5), 466–471. PMID:12847619

Alam, A. (2006). Knowledge of breast cancer and its risk and protective factors among women in Riyadh. *Annals of Saudi Medicine*, *26*(4), 272–277. PMID:16883082

Aldakheel, N. (2012). *Quality of primary healthcare in Saudi Arabia: a comprehensive review*.

Almalki, M. (2012). The relationship between quality of work life and turnover intention of primary healthcare nurses in Saudi Arabia. BMC Health Services Research, 12, 314 Almalki, M., Fitzgerald, G. & Clark M. Healthcare system in Saudi Arabia: An overview. *Eastern Mediterranean Health Journal*, *17*(10), 784–793. PMID:22256414

Altaf, F. (2006). Cervical cancer screening with pattern of pap smear. Review of multicenter studies. *Saudi Medical Journal*, *27*, 1498–1502. PMID:17013471

Anderson, B., Yip, Ch., Ramsey, S., Bengoa, R., Braun, S., Fitch, M., & Tsu, V. D. et al. (2006). Breast Cancer in Limited-Resource Countries: Healthcare Systems and Public Policy. *The Breast Journal*, *12*(s1), S54–S69. doi:10.1111/j.1075-122X.2006.00203.x PMID:16430399

Ayoola, E., & Gadour, M. (2004). Hepatocellular carcinoma in Saudi Arabia: Role of hepatitis B and C infection. *Journal of Gastroenterology and Hepatology*, *19*(6), 665–669. doi:10.1111/j.1440-1746.2003.03334.x PMID:15151622

Baghdadi, L. & Baghdadi, R. (2007). *Referral system from primary healthcare center*.

Beaglehole, R., Bonita, R., Horton, R., Adams, C., Alleyne, G., Asaria, P., & Watt, J. et al. (2011). Priority actions for the non-communicable disease crisis. *Lancet*, *10*, 46. PMID:21474174

Boyle, P., & Levin, B. (2008). *International Agency for Research on Cancer, 42*. [Electronic Version]. Retrieved 2015, from: http://www.iarc.fr/en/publications/pdfs-online/wcr/2008/wcr_2008.pdf

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Brown, A., Ravichandran, K., & Warnakulasuriya, S. (2006). The unequal burden related to the risk of oral cancer in the different regions of the Kingdom of Saudi Arabia. *Community Dental Health*, *23*(2), 101–106. PMID:16800366

Brown, D., Kjaer, S., Sigurdsson, K., Iversen, O., Hernandez-Avila, M., Wheeler, C., & Barr, E. et al. (2009). The Impact of Quadrivalent Human Papillomavirus (HPV; Types 6, 11, 16, and 18) L1 Virus-Like Particle Vaccine on Infection and Disease Due to Oncogenic Nonvaccine HPV Types in Generally HPV-Naive Women Aged 16–26 Years. *The Journal of Infectious Diseases*, *199*(7), 926–935. doi:10.1086/597307 PMID:19236279

Busse, R., Blümel, M., Kreinsen, D., & Zentner, A. (2010). *Tackling chronic disease in Europe Strategies, interventions and challenges, in Observatory Studies Series No 20. 2010*, World Health Organization, European Observatory on Health Systems and Policies. [Electronic Version]. Retrieved from: http://www.euro.who.int/__data/assets/pdf_file/0008/96632/E93736.pdf

Clifford, G. M., Goncalves, M. A. G., & Franceschi, S.HPV and HIV study group. (2006). Human papillomavirus types among women infected with HIV: A meta-analysis. *AIDS (London, England)*, *20*(18), 2337–2344. doi:10.1097/01.aids.0000253361.63578.14 PMID:17117020

CTR, A., CTR, O., Urcia, J. & Khan, A. (2012). *Tumor registry annual report 2011*. King Faisal Specialist Hospital and Research Centre. Retrieved from: http://www.kfshrc.edu.sa/oncology/Tumor%202011%20New%206%20Final.pdf

Dandash, Kh., & Al-Mohaimeed, A. (2007). Knowledge, Attitudes, and Practices Surrounding Breast Cancer and Screening in Female Teachers of Buraidah, Saudi Arabia. *International Journal of Health Sciences*, *1*(1), 61–71. PMID:21475453

Dayyani, F., Etzel, C., Liu, M., Ho, Ch., Lippman, S., & Tsao, A. (2010). Meta -analysis of the impact of human papillomavirus (HPV) on cancer risk and overall survival in head and neck squamous cell carcinomas (HNSCC). *Head & Neck Oncology*, *2*(1), 15. doi:10.1186/1758-3284-2-15 PMID:20587061

Dey, S., & Soliman, A. S. (2010). Cancer in the Global Health Era: Opportunities for the Middle East and Asia. *Asia-Pacific Journal of Public Health*, *22*(3), 75S–82S. doi:10.1177/1010539510372846 PMID:20566537

El-Attar, I. (2005). Cancer databases in the Arab world. *Ethnicity & Disease*, 15. PMID:15787030

El-Serag, H. (2007). *Epidemiology of Hepatitis C-Related Hepatocellular Carcinoma: HCV and HCC. Medscape Gastroenterology*. [Electronic Version]. Retrieved from: http://www.medscape.org/viewarticle/560012_3

Engelgau, M., Okamoto, K., Navaratne, K., & Gopalan, S. (2010). *Prevention and Control of Selected Chronic NCDs in Sri Lanka: Policy Options and Action. Health, Nutrition, and Population Family (HNP) of the World Bank's Human Development Network*. [Electronic Version]. Retrieved 2015, from: http://siteresources.worldbank.org/healthnutritionandpopulation/Resources/281627-1095698140167/NCDsSriLanka.pdf

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Evans, A., London, W., & Angstrom, P. (2005). *Liver Cancer and Hepatitis Viruses*. Cancer prevention. [Electronic Version]. Retrieved from: http://www.nypcancerprevention.com/issue/5/pro/pro_features/app_day.shtml

Farmer, P., Frenk, J., Knaul, F., Shulman, L., Alleyne, G., Armstrong, L., & Seffrin, J. et al. (2010). Expansion of cancer care and control in countries of low and middle income: A call to action. *Lancet*, *376*(9747), 1186–1193. doi:10.1016/S0140-6736(10)61152-X PMID:20709386

Gazette, S. (2012). *Economy: Saudi spending on healthcare to hit $24b in '12* [Electronic Version]. Retrieved from: http://www.saudigazette.com.sa/index.cfm?method=home.regcon&contentid=20121020140276

Geneau, R., Stuckler, D., Stachenko, S., McKee, M., Ebrahim, S., Basu, S., & Beaglehole, R. et al. (2010). Raising the priority of preventing chronic diseases: A political process. *Lancet*, *376*(9753), 1689–1698. doi:10.1016/S0140-6736(10)61414-6 PMID:21074260

Henry, S., Bosch, F., & Bowers, J. (2002). Aflatoxin, hepatitis and worldwide liver cancer risks. *Advances in Experimental Medicine and Biology*, *504*, 229–233. PMID:11922091

Hermine, O., Lefrère, F., Bronowicki, J., Mariette, X., Jondeau, K., Eclache-Saudreau, V., & Troussard, X. et al. (2002). Regression of Splenic Lymphoma with Villous Lymphocytes after Treatment of Hepatitis C Virus Infection. *The New England Journal of Medicine*, *347*(2), 89–94. doi:10.1056/NEJMoa013376 PMID:12110736

Hixon, A., & Maskarinec, G. (2008). The Declaration of Alma Ata on its 30th anniversary: Relevance for family medicine today. *Family Medicine*, *40*(8), 585–588. PMID:18988046

Hospedales, C., Samuels, T., Cummings, R., Gollop, G., & Greene, E. (2011). Raising the priority of chronic non-communicable diseases in the Caribbean. *Revista Panamericana de Salud Pública*, *30*(4), 393–400. PMID:22124699

Howard, J. J. (2014). Medical devices and the Middle East: market, regulation, and reimbursement in Gulf Cooperation Council states. Medical Devices (Auckland, NZ), 7, 385.

Ibrahim, E., Zeeneldin, A., El-Khodary, T., Al-Gahmi, A., & Bin Sadiq, B. (2008). (1). Past, Present and Future of Colorectal Cancer in the Kingdom of Saudi Arabia. *Saudi Journal of Gastroenterology*, *14*(4), 178–182. doi:10.4103/1319-3767.43275 PMID:19568534

Joseph, F., Perz, O., Armstrong, A., Farrington, L., Hutin, Y., & Bell, P. (2006). The contributions of hepatitis B virus and hepatitis C virus infections to cirrhosis and primary liver cancer worldwide. *Journal of Hepatology*, *45*(4), 529–538. doi:10.1016/j.jhep.2006.05.013 PMID:16879891

KFF. (2012). *Disparities in Health and Healthcare: Five Key Questions and Answers*. Available from: http://kff.org/disparities-policy/issue-brief/disparities-in-health-and-health-care-five-key-questions-and-answers/

Maghrabi, J., Al-Ghamdi, A. & Elhakeem, H. (2004). Pattern of skin cancer in Southwestern Saudi Arabia. *Saudi Medical Journal*, *25*(6), 776–779. PMID:15195210

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Marur, S., D'Souza, G., Westra, W., & Forastiere, A. (2010). HPV-associated head and neck cancer: A virus-related cancer epidemic. *The Lancet Oncology*, *11*(8), 781–789. doi:10.1016/S1470-2045(10)70017-6 PMID:20451455

Mazzaro, C., Tirelli, U., & Pozzato, G. (2005). Hepatitis C virus and non-Hodgkin's lymphoma 10 years later. *Digestive and Liver Disease*, *37*(4), 219–226. doi:10.1016/j.dld.2005.01.003 PMID:15788203

Mele, A., Pulsoni, A., Bianco, E., & Musto, P., Szklo, A., … Franceschi, S. (2003). Hepatitis C virus and B-cell non-Hodgkin lymphomas: An Italian multicenter case-control study. *Journal of American Society of Hematology*, *102*(3). PMID:12714514

Mellin, H., Friesland, S., Lewensohn, R., Dalianis, T., & Munck-Wikland, E. (2000). Human papillomavirus (HPV) DNA in tonsillar cancer: Clinical correlates, risk of relapse, and survival. *Int. J. Cancer, 89*, 300–304 [Electronic Version]. Retrieved from: http://onlinelibrary.wiley.com/doi/10.1002/1097-0215%2820000520%2989:3%3C300:AID-IJC14%3E3.0.CO;2-G/full

Midhet, F., Al Mohaimeed, A., & Sharaf, F. (2010). Dietary practices, physical activity and health education in qassim region of saudi arabia. *International Journal of Health Sciences*, *4*(1), 3–10. PMID:21475520

Mirza, T. (2006). *Quality of referral letters in AlRusaifah PHC Center in Makkah*. Saudi Society of Family and Community Medicine.

MOH. (2002). Principles and Practice of primary healthcare. *Health Education*, 101.

Mourshed, M., Hediger, V., & Lamber, T. T. (2007). *Gulf Cooperation Council Healthcare: Challenges and Opportunities*. Retrieved from: http://www.weforum.org/pdf/Global_Competitiveness_Reports/Reports/chapters/2_1.pdf

NCI, DHHS, & CDC. (2013). United States Cancer Statistics: 1999–2009 Incidence and Mortality Web-based Report. Atlanta, GA: Author.

Parkin, D. (2011). Cancers attributable to infection in the UK in 2010. *British Journal of Cancer*, *105*, S49–S56. doi:10.1038/bjc.2011.484 PMID:22158321

Pruitt, S., & Epping-Jordan, J. (2005). Preparing the 21st century global healthcare workforce. *BMJ (Clinical Research Ed.)*, *330*(7492), 637–639. doi:10.1136/bmj.330.7492.637 PMID:15774994

Qari, F. (2004). Pattern of thyroid malignancy at a University Hospital in Western Saudi Arabia. *Saudi Medical Journal*, *25*(7). PMID:15235690

RCJY. (2012). *Investors's Guide: The Future is Here*. [Electronic Version]. Retrieved from: http://www.rcjy.gov.sa/en-US/Yanbu/MediaCenter/Documents/Industrial%20Investor%27s%20Guide%202011.pdf

Sait, K., Bentley, J., Anfinan, N., & Power, P. (2012). Cervical Cancer Prevention in Saudi Arabia: It is Time to Call for Action! *The Open Women's Health Journal*, *6*(1), 1–5. doi:10.2174/1874291201206010001

Salim, E. I., Moore, M. A., Al-Lawati, J. A., Al-Sayyad, J., Bazawir, A., Bener, A., & Sobue, T. (2009). Cancer epidemiology and control in the arab world-past, present and future. *Asian Pacific Journal of Cancer Prevention*, *10*(1), 3–16. PMID:19469617

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Salim, E. I., Moore, M. A., Bener, A., Habib, O. S., Seif-Eldin, I. A., & Sobue, T. (2010). Cancer epidemiology in South-West Asia-past, present and future. *Asian Pacific Journal of Cancer Prevention*, *11*(Suppl 2), 33–48. PMID:20553067

Samb, B., Bolhuis, A., & Etienne, A. (2011). Delivering universal primary care services for non- communicable diseases: The need for stronger country health systems. Geneva: World Health Organization. [Electronic Version], Retrieved from http://apps.who.int/medicinedocs/documents/s18644en/s18644en.pdf

Santosa, H., & Rosario, N. (2008). Access to medication: Key to achieving treatment goals. *Bulletin of the World Health Organization*, *86*(6). PMID:18568261

Shulman, L., Willett, W., Sievers, A., & Knaul, F. (2010). Breast Cancer in Developing Countries: Opportunities for Improved Survival. *Journal of Oncology*, 2010. PMID:21253541

Steward, B., & Wild, C. (2014). *WHO World Cancer Report. International agency for research on Cancer/World Health Organization: Lyon, France*. [Electronic Version]. Retrieved from: http://www.iarc.fr/en/publications/books/wcr/

Taha, S., & Kamal, B. (2005). Screening program for prostate cancer at a university hospital in eastern Saudi Arabia. *Saudi Medical Journal*, *26*(7), 1104–1106. PMID:16047062

Tennant, B., Toshkov, I., Peek, S., Jacob, J., Menne, S., Hornbuckle, W., & Gerin, J. et al. (2004). Hepatocellular Carcinoma in the Woodchuck Model of Hepatitis B Virus Infection. *Gastroenterology*, *127*(5), S283–S293. doi:10.1053/j.gastro.2004.09.043 PMID:15508096

Unger, J., Paepe, P., & Green, A. (2003). A code of best practice for disease control programmes to avoid damaging healthcare services in developing countries. *The International Journal of Health Planning and Management*, *18*, S27–S39.

US Health News. (2012). *Millions of Healthy Years of Life Lost to Cancer Worldwide Health*. Retrieved from: https://odenvilledrugs.com/odenville/article.php?id=669617&category=-CER

Vedel, I. P. M., Monette, M., Monette, J., & Bergman, H. (2011). Barriers and facilitators to breast and colorectal cancer screening of older adults in primary care: A systematic review. *Urnal of Geriatric Oncology*, *2*(2), 85–98. doi:10.1016/j.jgo.2010.11.003

Vedsted, P., & Olesen, F. (2009). Early diagnosis of cancer-the role of general practice. *Scandinavian Journal of Primary Health Care*, *27*(4), 193–194. doi:10.3109/02813430903478623 PMID:19958062

WBO. (2012). Non-communicable diseases - South Asia: Opportunities for Prevention and Control. World Bank Organization.

WHO. (2003). *World Cancer Report*. Retrieved from: http://www.iarc.fr/en/publications/pdfs-online/wcr/2003/

WHO. (2008). *2008-2013 Action Plan for the Global Strategy for the Prevention and Control of Noncommunicable Diseases*. [Electronic Version]. Retrieved from: http://www.who.int/nmh/publications/ncd_action_plan_en.pdf

WHO. (2008). *Cancer Control WHO Guide for Effective Programmes*. WHO.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

WHO. (2008). *Commission on the Social Determinants of health. World Health Organization: Geneva*. [Electronic Version]. Retrieved 2015, from: http://www.who.int/social_determinants/thecommission/finalreport/en/

WHO. (2011). *Action plan for implementation of the European Strategy for the Prevention and Control of Noncommunicable Diseases 2012–2016*. World Health Organization. [Electronic Version]. Retrieved from: http://www.euro.who.int/__data/assets/pdf_file/0019/170155/e96638.pdf

WHO. (2012). *Hepatitis B; Fact sheet N°204*. World Health organization. [Electronic Version]. Retrieved from: http://www.who.int/mediacentre/factsheets/fs204/en/

WHO. (2013). *Cancer: Fact sheet N°297*. World Health Organization Report. [Electronic Version]. Retrieved from: www.who.int/mediacentre/factsheets/fs297/en/

WHO/PHAC. (2005). *Preventing chronic diseases: a vital investment: WHO global report*. [Electronic Version]. Retrieved 2015, from: http://www.who.int/chp/chronic_disease_report/part1/en/

Wong, C., & Goh, K. (2006). Chronic hepatitis B infection and liver cancer. *Biomedical Imaging and Intervention Journal*, *2*(3), e7. doi:10.2349/biij.2.3.e7 PMID:21614253

Yusuf, S., Hawken, S., Ounpuu, S., Dans, T., Avezum, A., Lanas, F., & Lisheng, L. et al. (2004). Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the internet study): Case-control study. *Lancet*, *364*(9438), 937–952. doi:10.1016/S0140-6736(04)17018-9 PMID:15364185

This work was previously published in Transforming Public Health in Developing Nations edited by Mohamud Sheikh, Aziza Mahamoud, and Mowafa Househ, pages 52-67, copyright year 2015 by Information Science Reference (an imprint of IGI Global).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 25
My Desired Self, Avatar:
The Impact of Avatar Creation on Persuasion

Youjeong Kim
New York Institute of Technology, USA

ABSTRACT

In computer-mediated communication (CMC) environments, users utilize their avatars as a communication channel to interact and connect with others, and they choose and create them accordingly to represent their self. As such, several major question areas arise: 1) As an extension of identity, how does a user customize his/her avatar? How is the avatar's appearance related to the avatar creator's self-concept? 2) How does avatar creation influence the avatar creator's psychological and behavioral consequences? To answer these questions, the current study leveraged a Korean social networking site, which currently provides avatars called "Minimis," in a randomized experimental setting. This study found that the more the participants perceived their avatars to look like their desired selves, the more likely they evaluated their avatars as being attractive, credible, confident, cool, capable, and persuasive, but failed to find a significant relationship between avatar users' perceptions toward self-created avatars and their attitudes toward the social network site or ads.. The limitations and implications will be discussed.

1. INTRODUCTION

With the development of communication technology, it is remarkable that, unlike face-to-face communication, people are able to construct their identities in cyberspace. The technology enables people to create who they are. In particular, the avatar, as an extension of identity, has recently been focused on by many scholars in communications as well as online psychology.

An avatar is a graphical icon that represents users in online communities, chat rooms, or virtual reality games (Nowak, 2000). An avatar is not simply an embodied graphic that allows users to look at their creatures from a third-person perspective (Taylor, 2002) but is a visualized identification and a virtual self that is imbued with a user's feeling of presence in the virtual space. It allows users to customize and change the appearance of everything from skin color to eye size and even gender easily.

DOI: 10.4018/978-1-5225-0549-5.ch025

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

In terms of power of anonymity, avatars motivate users to activate one of their inner selves. Through avatars, users present themselves either in their actual forms that are present in the public or in ideal forms that have not been shown in public but are what the users desire to be. Throughout the study, we attempt to explicate the sense of self that could be expressed through avatar creation and its persuasive role in influencing avatar users' perceptions and attitudes.

To sum up, the current study aims to answer the following questions: 1) as the extension of identity, how do people customize their avatars? How does the avatar look like the user? Does it look like the user's actual self or their desired self? Are there any perceptional differences between the actual avatar and the desired avatar? 2) If they are different, how do users perceive the desired avatar? 3) If users perceive them positively, what role does the perception of the desired avatar play in cyberspace, and what are the practical implications of it?

The answers will significantly contribute to the communication literature by successfully bridging the missing link between the avatar as an extension of self and as an effective communication tool in the context of persuasion.

2. LITERATURE REVIEW

2.1. Self-Identity and Avatars

It is not a novel idea that a person possesses multiple personas, also known as *identities* in the field of psychology (Goffman, 1959; Jung, 1953). Goffman broadly classified a person's self with public self and inner self. Based on Jung's conceptualization, the inner self is one's real individuality, which resides in the unconscious self. Carl Rogers (1951) called this the *true self*. According to Rogers' argument, the true self is the one that psychologically exists in the present but is not fully expressed in social life. The true self is more often and easily expressed through the Internet than in traditional, face-to-face communication. Along these same lines, Bargh, McKenna, and Fitzsimons (2002) also provided reasonable explanations, arguing that the Internet enables people to facilitate self-expression through two powerful features: anonymity and freedom from self-disclosing negative aspects of oneself. As the analogy of 'strangers on a train' (Rubin, 1975) implies, the Internet is like a train in which people open up and disclose their inner or true selves to strangers sitting in the next seat.

Another powerful feature of the Internet that Bargh et al. argue is the absence of a physical "gating feature" (p.44). This is an extension of the power of anonymity, which gives users freedom to manipulate their identities. Invisibility in cyberspace allows users to be anybody and activate any one of their identities without necessarily developing a new or strange identity. Although the figure or appearance of the online and offline identities may look different, they are somewhat connected (Jordan, 1999; Kang & Yang, 2006; MacKinnon, 1995).

Given that, an avatar is a graphical feature that conveys one's online identity (Jordan, 1999). It represents users in online communities, chat rooms, or virtual reality games (Nowak, 2004). Suler (1999) refers to an avatar as "a picture, drawing, or icon that users choose to represent themselves." First used in Neal Stephenson's (1993) science fiction book, *Snow Crash* (Vilhjalmsson, 1997), the word "avatar" was originally comprised of two Sanskrit words, "Ava" (meaning "descend" or "pass") and "Terr" (meaning "beneath" or "earth"). Therefore, avatar literarily, means a "descent" into a lower realm of existence. In

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Hindu philosophy, avatar is used to refer to the incarnation (bodily manifestation) of a divine being on Earth (Lee, 2003). Although avatars in Internet environments are only used to represent the visualized embodiment of a character, the broader meaning of the ancient Indian philosophy is embedded in the concept of the avatar. In terms of the embodied character, some researchers have used the term "avatar" with the term "agent" interchangeably. However, Nowak (2000) distinguishes the avatar from the agent, saying that the latter is controlled by a computer program while the avatar is controlled by a user. An attractive feature of avatars is that they allow people to customize their appearance by selecting body parts and accessories from skin color to clothes and change it as often as they want. In other words, people are easily able to change and manipulate the appearance of avatars. It could be either identical to or totally different from the appearance of the real person. However, as discussed above, an avatar is not totally disassociated with user's self but reflects his or her identity to some extent.

2.2. Actual vs. Desired Self

The understanding of how people customize their avatars can provide insight into users' sense of self in cyberspace. We can easily find people who create avatars very similar to their real appearances, and some create avatars very different from themselves. Higgins (1987) refers to the first as the *actual self.* According to him, the actual self pertains to the appearance or characteristics that one possesses and shares with others. Regarding the self that is different from the real self, Carl Rogers (1951) calls this the *true self,* and Higgins (1987) calls it the *ideal self.* Both of these concepts share the common idea that the presented self is evidently the extension of identities that are not fully expressed in reality.

However, we found that these concepts do not adequately overarch the sense of self generated in the process of creating an avatar, particularly focusing on its appearance. In particular, the true self emphasizes the personalities or traits that people actually possess but are placed in one's sub-consciousness. Higgins' ideal self is accompanied with the concept of an individual's future self, which is the self that he/she could potentially possess and could someday develop but is not currently present.

This issue calls for the need to develop an adequate term to define an individual's self that appears different from the individual in reality but represents the appearance he or she desires. We, therefore, call this type of self the *desired self.* This term is distinguished from Rogers's true self or Higgins' ideal self in terms of current desire regardless of what the individual actually possesses. The desired self does not imply the creator's potential to become the created self, and it is not a future version of the self. Moreover, it can explicitly appear in the process of creating an avatar.

In short, the actual avatar is the one that is created to be identical or similar to one's real self, and the desired avatar is the one that appears different from the creator's real self but represents what the creator wishes to be.

3. THEORETICAL BACKGROUND

The conceptualization of the self through avatars raises some theoretical questions: 1) why do people differ in their self-descriptions when creating avatars, and 2) why is avatar creation meaningful to study in the field of communication? The answers to these questions might help us understand the persuasive effects avatars have, which is what the current study aims to investigate.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

3.1. Why Do People Differ in Their Self-Descriptions When Creating Avatars?

Snyder (1987) asserts that people "monitor the public appearance of self they display in social situations" (p.4). Presumably, avatar creation as a part of building identity would be associated with one's self-monitoring to some extent. According to Snyder's conceptualization, people are divided into two groups depending on how much they monitor themselves: high self-monitors and low self-monitors. High self-monitors are more likely to be concerned about others and to easily change their public identities based on situational cues than low self-monitors. Given that, it appears that high self- monitors are more likely to create desired avatars, and low-monitors are more likely to create actual avatars.

In addition to users' self-concept, Kang and Yang (2006) found that the characteristics of computer-mediate communication (CMC) media influence the choice of avatars. CMC provides synchronous interactive communication in which people communicate with others who have never met (e.g., Internet-based Relay Chat, or IRC) or are in a close relationship (e.g., peer-to-peer based instant messenger, or IM). Kang and Yang found that IM users were more likely to disclose themselves, and the similarities of the appearances between the real person and his/her chosen avatar was much closer as compared to IRC users. The study illustrates the anonymity of the Internet that was discussed above. Apparently, IM environments in which people in close relationships are mainly communicating with one another has a lower level of anonymity than IRC environments. Given that, the level of anonymity appears to play a key role in self-disclosure and, in turn, influences avatar creation.

3.2. Then, Why Is Avatar Creation Meaningful to Study in the Field of Communication?

So far, we have spent several sections mostly discussing the idea that what we wear reflects who we are. This is true; however, it is also true that what we wear determines who we are and how we behave. Among the numerous studies that have attempted to reveal the relationship between what people wear and how they behave, the study conducted by Frank and Gilovich (1988) found that subjects who simply wear black uniforms behaved more aggressively. A study replicated in a digital environment also confirmed consistent patterns by showing that users given avatars in a black robe expressed more antisocial behaviors than users given avatars in a white robe (Merola, Penas, & Hancock, 2006). Beyond these antisocial behavioral patterns, the effect of deindividuation (or the SIDE effect, the social identity model of deindividuation effect) emphasizes behavioral change depending on anonymity or local group norms (Postmes, Spears, & Lea, 1998, 2000; Spears & Lea, 1994).

However, the point that the current study aims to highlight is the behavioral conformity related to individual identity cues, particularly through avatar creation, which is consistent with Yee and Bailenson's (2007) study. Recently, Yee and Bailenson explored how the avatar changed the avatar user's behavior in cyberspace, believing that "transforming appearance permeates the culture" (p.271). Theorizing the proteus effect, Yee and Bailenson argue that "being an attractive [person]" and "being in a group of attractive people" is conceptually different, and they emphasize the individual but not an individual in a group. Their study theoretically supports their findings that when given an attractive avatar, the user behaved more friendly and socially.

These results invoke the feasibility of avatar use as a clinical or marketing tool for persuading avatar users. Another of Yee and Bailenson's (2006) studies proved that the experience of embodying one's

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

self through an avatar is influential when making a judgment on and having attitudes about others, thereby reducing negative stereotypes. The current study investigates the impact of avatar creation as a persuasive marketing tool.

3.3. Research Hypotheses

It is true that everybody has a desire to be attractive and credible, although the degree might be different depending on age and gender. However, attractiveness and credibility are self-judgmental terms that are perceived differently by individuals. Therefore, the perception of one's real self determines the distance between one's actual self and one's desired self. For example, if person A perceives himself/herself as unattractive, the gap between one's actual self and one's desired self might be large, even if everybody else considers person A to be attractive enough. Presumably, therefore, there might be a judgmental difference between the avatar that looks like the person's actual self and the avatar that looks like the person's desired self.

H1: There would be a difference in the judgment between the avatars that actually look like the person and the avatars that the person wishes to look like.

Apparently, many believe that one's desired self is more attractive and/or credible than one's actual self. In another line of research, Langlois and his colleagues (2000) found that attractive people behave more positively, exhibiting high levels of self-confidence, good mental health, and favorable attitudes. Based on their findings, it is predicted that an avatar like one's desired self is perceived to be more attractive and more credible, and, consequently, the desired-avatar users will show more favorable attitudes and behaviors than actual-avatar users. In sum, as shown in Yee and Bailenson's (2007) study, the type of avatar created for representing the real self influences the avatar user's behavior or attitudes. Therefore, it leads to the following hypotheses:

H2: When perceiving the avatars created to be like their desired selves, participants will evaluate the avatar as more attractive, credible, confident, cool, capable, and/or persuasive as compared to avatars created to be like their actual selves.

H3: When perceiving the avatars created to be like their desired selves, participants will evaluate the website in which the avatar appeared and the ad and PSA that are embedded in the website more favorably as compared to when they viewed these aspects using avatars created to be like their actual selves.

4. METHODS

4.1. Study Design

To explore the impact of avatars created for representing a user's self on a website, a within-subjects design was employed. On the manipulated website, participants were directed to create avatars that would represent the user on the website. They were then asked to browse the website in which the created avatar appeared.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

4.2. Participants

Twenty-two undergraduate students (9 men and 13 women) at a major northeast university participated in the experiment for extra course credit.

4.3. Stimulus

For the ecologically valid study, we searched for websites that currently provide an avatar-creating service and allow users to customize their avatars. Furthermore, to rule out the confounding effect of CMC media characteristics (Kang & Yang, 2006) and the familiarity of the website, we were very careful in selecting the website. After the rigorous examination, we finally manipulated an online Korean community site, Cyworld.com (http://www.cyworld.com), which has launched in the U.S. recently. Similar to Facebook.com or Myspace.com, Cyworld has a personal mini-homepage in which users are able to post pictures and share them with friends easily. In addition, this website provides avatars called "Minimis" for representing the user. Once a user creates the avatar, it appears on the left side of the front page. In addition, for the current study, we created two interactive square ads on the right side of the front page: one is a shampoo commercial and the other is a skin cancer PSA. Both ads are identical in size and structure to rule out the effect of ad size or order (see figure 1).

4.4. Procedures

The experiment was administered to participants in a computer lab. As participants arrived at the lab, they were greeted by the experimenter. The experimenter told the participants to have a seat in front of the monitor, asked them to read and sign the informed consent form, and directed them to study the instructions on the computer screen, which provided general instructions for the overall study. After making sure that participants understood the procedures described in the instructions, they were asked to watch a three-minute introductory video clip on the computer screen by clicking the button saying "click here to go to the next page." The video clip contains the introduction to Cyworld.com and the general instructions for creating avatars including how to customize them and how they function. After watching the video, participants were instructed to create an avatar, which would represent them in Cyworld while they were browsing. For the customization, four options were provided: skin color, hair styles, clothing tops, and clothing bottoms (see figure 2). Participants were able to observe the changes in their avatars as they selected one of the options and were allowed to play with them as much as they wanted. Then, if participants were satisfied with the avatars they created, after saving them, they could go to the next section.

Before browsing Cyworld, participants were asked to fill out online questionnaires for judging the avatars that they created in the previous section. Then, every participant was asked to browse the manipulated Cyworld, in which all of the links had been disabled for the study, and the two interactive ads for the shampoo and skin cancer had been added, for one minute and evaluate what they saw. When browsing the website, participants were able to observe the avatars they created in the previous section on the left side of the front page. After one minute, a pop-up window was scheduled to appear that led participants to the questionnaire.

To make sure that study participants had paid attention to the ads, participants were asked to write everything that they could recall about the ads and were asked to find the right answer among multiple-

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 1. Example of a website with an avatar-creating service



choice questions about the ads in order to measure participant recognition. Lastly, participants were asked to fill out online questionnaires that included evaluation criteria for the website, its ad, and its PSA, as well as demographic information.

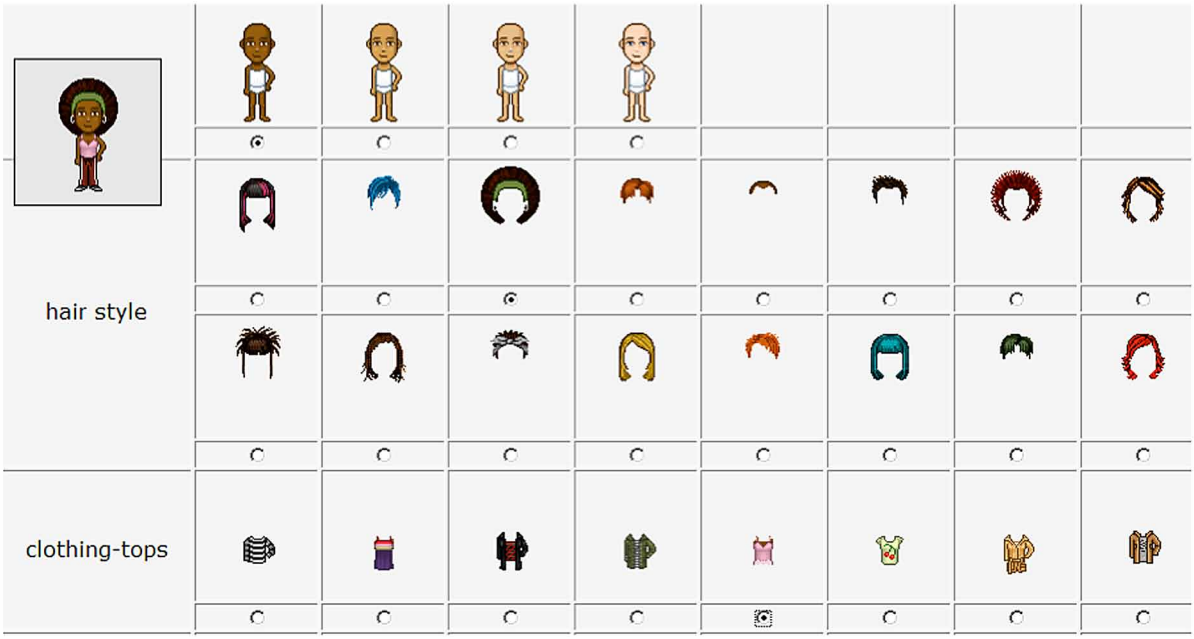
4.5. Variables

4.5.1. Independent Variable

The current study aims to investigate if 1) there is a judgmental difference between the actual avatar and the desired avatar, 2) if so, how users perceive their desired avatar in terms of positive attributes, and lastly, 3) if positively perceived, how it affects desired avatar users' attitudes toward the website or the ads embedded in it. To investigate the first issue, the participants' perceptions toward the avatar they customized was assessed. Participants were asked to rate the following statements on a 1-10 scale, with 1 representing *strongly disagree* and 10 representing *strongly agree*: Two items for the actual avatar ("My avatar looks like me" and "My avatar is similar to me") and two items for the desired avatar ("My avatar looks like what I want to be" and "My avatar represents what I would ideally want to be").

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 2. The manipulation of avatar customization



4.5.2. Dependent Variables

To address the second issue about the perceived attributes of the desired avatars, six 10-point semantic differential scales were used: attractive, credible, confident, cool, capable, and persuasive. Items were anchored between *not at all* (1) and *very much* (10). And, regarding the last issue about the avatar users' attitudes toward the website and ads, several items were adapted from literature. First, for the website, nine 10-point Likert-type items were adapted from Sundar and Kalyanaraman's (2004) website perceptions. Participants were asked to indicate how well each of the following adjectives described the Cyworld website: appealing, likeable, entertaining, interesting, attractive, stimulating, trustworthy, honest, and believable. Items were anchored between *very poorly* and *very well*.

Secondly, in order to measure users' attitudes toward the ads, ten 10-point semantic differential scales were used (Witte, 2005): stimulating, interesting, persuasive, informative, relevant to me, bad/good, appealing, believable, important, and attractive. In addition to these ten items, participants were also asked to rate the following questions on a 1-10 scale, with 1 representing *not at all* and 10 representing *a lot*: "How anxious did the shampoo ad make you feel about your hair?" and "How uncomfortable did the shampoo ad make you feel about your hair?"

Lastly, users' attitudes toward the skin cancer PSA were measured with the same items that were used for the ad (Witte, 2005). Ten 10-semantic differential scales were employed (stimulating, interesting, persuasive, informative, relevant to me, bad/good, appealing, believable, important, and attractive), and users were asked the same two questions on a 1-10 scales, with 1 representing *not at all* and 10 representing *a lot*: "How anxious did the 'Don't Toast' skin cancer PSA make you feel about your hair?" and "How uncomfortable did the 'Don't Toast' skin cancer PSA make you feel about your hair?" Three additional questions asked participants to rate their agreement to the following statements: "skin cancer is

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

a severe health problem," "skin cancer is a serious threat to my health," and "skin cancer is a significant disease." These items were measured on a 1-10 scale, ranging from *strongly disagree* to *strongly agree*.

4.6. Index Construction and Preparation for Data Analysis

An exploratory factor analysis was conducted upon the ten items relating to participants' attitudes toward the website, yielding two factors accounting for 89 percent of the variance. Pursuant to rotation, seven items (entertaining, appealing, interesting, stimulating, likeable, attractive, and intellectual) were loaded under the first factor labeled "attitude toward the website". This item exhibited a high degree of internal consistency (Cronbach's $\alpha = .97$). Three items (trust, honest, believable) were loaded under the second factor labeled "cognitive perception of the website" (Cronbach's $\alpha = .93$).

For attitudes toward the ad, twelve items yielded two factors, accounting for 72 percent of the variance. After rotating, five items (persuasive, informative, relavant, believable, good) were loaded under the first factor labeled "attitude toward the ad" (Cronbach's $\alpha = .837$). The other three items (degree to be anxious and uncomfortable, and important) were labeled "worry about hair" (Cronbach's $\alpha = .751$). The rest of items (stimulating, appealing, interesting, and attractive) were dropped due to the cross-loading.

The fifteen items relating to attitudes toward the PSA yielded three factors, accounting for 88 percent of the variance. After rotating, seven items (interesting, stimulating, persuasive, good, appealing, believable, and relevant) were loaded under the first factor labeled "attitudes toward the PSA" (Cronbach's $\alpha = .97$). Three items (skin cancer is a severe health problem, skin cancer is a serious threat to my health, and skin cancer is a significant disease) were loaded under the second factor labeled "attitude toward skin cancer" (Cronbach's $\alpha = .92$). The last two items (degree to be anxious and uncomfortable) were labeled "worry about skin cancer" (Pearson's $r = .93$, $p < .001$).

5. RESULTS

To examine the participants' judgmental differences between the actual and desired avatars, a one-way repeated measures ANOVA employing a multivariate approach was conducted. Consistent with H1, this analysis showed that there was a marginally significant difference in perceiving the self-created avatar, Wilks' $\Lambda = .85$, $F(1, 21) = 3.74$, $p = .06$, partial $\eta = .15$ (See table 1).

To test H2 addressing the judgmental issue about the desired avatar's positive attributes, six separate linear regression tests were performed. This analysis showed that when judging the self-created avatar to be like one's desired self, users perceived their avatar as being more attractive ($\beta = .39$, $t = 2.25$), credible ($\beta = .42$, $t = 2.35$), confident ($\beta = .43$, $t = 2.53$), cool ($\beta = .454$, $t = 2.86$), capable ($\beta = .40$, $t =$

Table 1. Perceived self-resemblance of self-created avatars

	Avatar Type	
	Actual Avatar	**Desired Avatar**
M	6.023	4.955
SE	.449	.498

Wilks' $\Lambda = .85$, $F(1, 21) = 3.74$, $p = .06$, partial $\eta = .15$

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

2.53), and persuasive (β =.40, t = 2.53) (see table 2). On the other hand, the analysis revealed that the perception of the actual avatar was marginally associated with the attractiveness among the aforementioned attributions. In other words, the more participants perceived their created avatars to be like their actual selves, the more they perceived their avatars as being attractive (β =.352, t = 2.749). Therefore, H2 was supported.

H3 predicted that the desired avatar users would evaluate the website in which the desired avatar appeared and ads that were embedded in the website. To test this, another set of regression tests was performed. The results failed to illustrate significant effects of avatar perception in predicting positive attitudes toward the website, ad, and PSA. Therefore, H3 was not supported.

6. DISCUSSION

In general, the current study has attempted to answer the following questions: 1) as the extension of identity, how do people customize their avatars? In what ways do the avatars look like users? Do they look like the users' actual selves or their desired selves? Is there any perceptional difference between the actual avatar and the desired avatar? 2) If they are different, how do users perceive the desired avatar? 3) If users positively perceive their avatar, what role does the perception of the desired avatar play in cyberspace, and, what are the practical implications of this?

In regard to the first issue, we found that there was a marginally significant difference in evaluating the self-customized avatar, particularly judging if the avatar looked like the avatar creator's actual image or their desired image. Furthermore, the current study revealed that participants were more likely to create an actual avatar that resembled the avatar creator's real image rather than their desired image for representing themselves in the social networking site (SNS). As an extension of Kang and Yang's (2006) study, which revealed that avatar choice differed by the characteristic of CMC media and subsequently different levels of anonymity, this study found that the actual avatar was preferred over the desired avatar by the users of the SNS in which the level of anonymity is mixed and flexible (e.g., when used for maintaining existing relationships with close friends, it is considered to have low anonymity, and when used for developing new relationships with unknown people in the network, it is considered to have high anonymity).

Table 2. Regression analysis for avatar perception

	Desired avatar		**Actual avatar**	
	Beta	**t**	**Beta**	**t**
Attractive	.39**	2.25	.35*	1.75
Credible	.42**	2.35	.19	.88
Confident	.43**	2.53	.20	.96
Cool	.45**	2.86	.26	1.30
Capable	.40**	2.53	.23	1.18
persuasive	.40**	2.53	.23	1.18

** $p < .05$, * $p < .1$

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

In addition, this study answered the question as to how the desired avatar was evaluated when users perceived the avatar as a being like their desired selves. Consistent with Langlois et al.'s (2000) study, the desired avatar users perceived their avatar to be attractive, credible, confident, cool, capable, and persuasive. Furthermore, interestingly, the results indicated that the users' actual avatar perceptions were marginally associated with their perceptions of attractiveness. Regarding this, Pelham and his colleagues (2003) provided a good theoretical argument of "implicit egotism" for explaining the positive attitudes toward self-related things. A marketing research study (Brendl, Chattopadhyay, Pelham, and Carvallo, 2005) found that people gravitated toward things that resemble the self and, in turn, evaluated them positively. Given that, it makes sense that avatar users consider the self-resembling avatar to be attractive regardless of whether it looks like the user's actual image or his/her desired image.

For the next step, this study expected that the positive valence toward the self-created avatar, which particularly resembles the avatar creators' desired image, would transfer to the website in which the avatar appeared or to the web contents, such as an ad or a PSA, and would generate positive attitudes. This expectation was based on the theoretical framework of the "attribute-specific valence transfer" (Brendl, Chattopadhyay, Pelham, and Carvallo, 2005) in which self-related contents might generate overall liking for the ad or PSA. Unfortunately, this study failed to find any persuasive effects of the website, its ad, or its PSA.

The current study notes several limitations in the study design. The first critical problem was that the options and choices for customizing avatars were relatively small and un-diverse since we randomly chose from the existing avatar options. The desired image that people want usually includes various appearances from body size to hairstyle. However, this study only provided choices for skin color, hairstyle, and clothing. It may have made it difficult for people to draw a line between the actual avatar and the desired avatar and, in turn, may have influenced participants' perceptions toward their avatar and the website. Secondly, for an ecologically valid study and, at the same time, to reduce the compounding effect of familiarity, we used an existing Korean social networking site (SNS) that provides the avatar service. However, it generated the ecological validity issue since most participants in our study were only familiar with avatars from virtual reality games or chat rooms, but not from SNS. In addition, the disability of the links for the purpose of this study made participants feel uncomfortable, leading to the low favorability of attitudes regarding the website, compared to attitudes toward the ad or the PSA. The most significant pitfall of the current study is the limited role of the avatar in the SNS. In the chatting environment or a normal SNS, the avatars appear next to users' writings. However, for the purpose of this study, we just briefly touched on the function of avatars in the introductory video clip. While browsing the website, participants merely saw their self-created avatar on the left side of the webpage. It might have hampered how users idenified with their self-created avatars.

These limitations suggest ideas for future studies. More importantly, it is necessary to employ the various options and choices for customizing avatars. Some components, such as body size (including height or weight) are not separable to determine attractiveness. Furthermore, to explore the persuasive effects of avatars, a future study should investigate an environment of avatar-mediated communication. In addition, with the development of new technology, it is becoming more common to see the 3D avatars in cyberspace. Therefore, this could be a considered in future studies.

This study has explicated the avatar as an extension of identity and has shed light on the persuasive effect of avatars in the SNS context. The advance of new technology was initially blamed for de-individualizing and unitizing people under the label of "mass." Given that, the avatar enables people to focus more on the self and to *see* their existence through a self-resembling object. As we reviewed in

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

this study, people have positive valence toward self-resembled objects. In particular, creating an avatar is more meaningful beyond just being a self-resembling creature. It is a process of building identity and thereby increasing self-awareness and self-preservation. Consequently, it may help with designing effective health messages focusing on the self.

In addition, through creating avatars, people make their desired selves that are hidden under their inner selves real and visible. We believe that this kind of avatar creation might help some people who have suffered mentiall by ought self which is obligated to be in real life. Furthermore, as Yee and Bailenson (2007) and this study revealed, when taking on an attractive avatar (identity), participants were more likely to present positive attitudes and behaviors. This psychologically supports the idea that people who are judged to be attractive by themselves and/or by others show high levels of confidence, which increase their self-esteem (Langlois et al., 2000). Although this study could not find any powerful effects of users' perceptions of their self-created avatars on their evaluation of the website, ad, and PSA, given that the virtual population, including Internet users, has grown epidemically, this study will contribute to the development of more effective intervention programs as well as more effective advertising on websites.

REFERENCES

Bargh, J. A., McKenna, K. Y. A., & Fitzsimons, G. M. (2002). Can you see the real me? Activation and expression of the "true self" on the Internet. *The Journal of Social Issues*, *58*(1), 33–48. doi:10.1111/1540-4560.00247

Brendl, C. M., Chattopadhyay, A., Pelham, B. W., & Carvallo, M. (2005). Name letter branding: Valence transfers when product specific needs are active. *The Journal of Consumer Research*, *32*(3), 405–415. doi:10.1086/497552

Frank, M. G., & Gilovich, T. (1988). The dark side of self-and social perception: Black uniforms and aggression in professional sports. *Journal of Personality and Social Psychology*, *54*(1), 74–85. doi:10.1037/0022-3514.54.1.74 PMID:3346809

Goffman, E. (1959). *The Presentation of Self in Everyday Life*. New York: Doubleday.

Higgins, E. T. (1987). Self-discrepancy theory. *Psychological Review*, *94*(1), 120–121. PMID:3615707

Jordan, T. (1999). *Cyberpower: The culture and politics of cyberspace and the internet*. New York: Routledge. doi:10.4324/9780203448632

Jung, C. C. (1953). *Two essays in analytical psychology* (R. F. Hull, Trans.). New York: Pantheon.

Kang, H. S., & Yang, H. D. (2006). The visual characteristics of avatars in computer-mediated communication: Comparison of internet relay chat and instant messenger as of 2003. *International Journal of Human-Computer Studies*, *64*(12), 1173–1183. doi:10.1016/j.ijhcs.2006.07.003

Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin*, *126*(3), 390–423. doi:10.1037/0033-2909.126.3.390 PMID:10825783

MacKinnon, R. (1995). Searching for the leviathan in usenet. In S. Jones (Ed.), *Cybersociety* (pp. 112–138).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Merola, N., Penas, J., & Hancock, J. (2006). *Avatar color and social identity effects: On attitudes and group dynamics in virtual realities.* Paper presented at the ICA 2006, Dresden, Germany.

Nowak, K. L. (2004). The Influence of anthropomorphism and agency on social judgment in virtual environments. *Journal of Computer-Mediated Communication*, *9*(2), 1–28.

Postmes, T., Spears, R., & Lea, M. (1998). Breaching or building social boundaries? SIDE-effects of computer-mediated communication. *Communication Research*, *25*(6), 689–715. doi:10.1177/009365098025006006

Postmes, T., Spears, R., & Lea, M. (2000). The formation of group norms in computer-mediated communication. *Human Communication Research*, *26*(3), 341–371. doi:10.1111/j.1468-2958.2000.tb00761.x

Rogers, C. R. (1951). *Client-centered therapy*. Boston: Houghton-Mifflin.

Rubin, Z. (1975). Disclosing oneself to a stranger: Reciprocity and its limits. *Journal of Experimental Social Psychology*, *11*(3), 233–260. doi:10.1016/S0022-1031(75)80025-4

Snyder, M. (1987). *Public appearances, private Realities: The psychology of self-monitoring*. W H Freeman.

Spears, R., & Lea, M. (1994). Panacea or panopticon? The hidden power in computer-mediated communication. *Communication Research*, *21*(4), 427–459. doi:10.1177/009365094021004001

Suler, J. (1999). *The psychology of cyberspace*. Rider University.

Taylor, T. L. (2002). Living digitally: Embodiment in virtual worlds. In R. Schroeder (Ed.), *The Social Life of Avatars: Presence and Interaction in Shared Virtual Environments*. Springer. doi:10.1007/978-1-4471-0277-9_3

Vilhjalmsson, H. H. (1997). *Autonomous Communicative Behaviors in Avatars*. Massachusetts Institute of Technology.

Yee, N., & Bailenson, J. N. (2006). *Walk a mile in digital shoes: The impact of embodied perspective-taking on the reduction of negative stereotyping in immersive virtual environments.* Paper presented at the 9th Annual International Workshop on Presence, Cleveland, Ohio.

Yee, N., & Bailenson, J. N. (2007). The proteus effect: The effect of transformed self-representation on behavior. *Human Communication Research*, *33*(3), 271–290. doi:10.1111/j.1468-2958.2007.00299.x

This work was previously published in the International Journal of Virtual Communities and Social Networking (IJVCSN), 7(1); edited by Subhasish Dasgupta, pages 1-13, copyright year 2015 by IGI Publishing (an imprint of IGI Global).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 26
Isocenter Verification in Radiotherapy Clinical Practice Using Virtual Simulation:
An Image Registration Approach

George Matsopoulos
National Technical University of Athens, Greece

Vasiliki Markaki
National Technical University of Athens, Greece

Pantelis A. Asvestas
Technological Educational Institute of Athens, Greece

Kalliopi Platoni
University of Athens, Greece

Vasilios Kouloulias
University of Athens, Greece

ABSTRACT

This chapter presents an overview of the procedures that are used for the verification of the patient position during radiotherapy. Furthermore, a method for the verification of the radiotherapy isocenter prior to treatment delivery is proposed. The method is based on the alignment of two Computed Tomography (CT) scans: a scan, which is acquired for treatment planning, and an additional verification scan, which is acquired prior to the treatment delivery. The proposed method was applied to CT scans, acquired from 20 patients with abdominal tumors and 20 patients with breast/lung cancer. The results of the proposed method were compared with the ones obtained using conventional methods, indicating that the estimated isocenter displacement can be translated into patient setup error inside the treatment room.

INTRODUCTION

According to the World Health Organization (WHO), chronic diseases (also known as not noncommunicable diseases) are long duration diseases that are not passed from person to person (World Health Organization, 2014). The main types of chronic diseases are cardiovascular diseases (like heart attacks and stroke), cancers, respiratory diseases (such as chronic obstructed pulmonary disease and asthma)

DOI: 10.4018/978-1-5225-0549-5.ch026

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

and diabetes. Cancers can be considered the most serious chronic diseases with significant impact on the patients. According to Ferlay et al. (2013), "*there were 14.1 million new cancer cases, 8.2 million cancer deaths and 32.6 million people living with cancer (within 5 years of diagnosis) in 2012 worldwide. 57% (8 million) of new cancer cases, 65% (5.3 million) of the cancer deaths and 48% (15.6 million) of the 5-year prevalent cancer cases occurred in the less developed regions. The overall age standardized cancer incidence rate is almost 25% higher in men than in women, with rates of 205 and 165 per 100,000, respectively. Male incidence rates vary almost five-fold across the different regions of the world, with rates ranging from 79 per 100,000 in Western Africa to 365 per 100,000 in Australia/New Zealand (with high rates of prostate cancer representing a significant driver of the latter). There is less variation in female incidence rates (almost three-fold) with rates ranging from 103 per 100,000 in South-Central Asia to 295 per 100,000 in Northern America. In terms of mortality, there is less regional variability than for incidence, the rates being 15% higher in more developed than in less developed regions in men, and 8% higher in women*".

One of the most common methods for cancer treatment is radiation therapy, which *is a clinical modality dealing with the use of ionizing radiations in the treatment of patients with malignant neoplasias (and occasionally benign diseases)* (Halperin, Brady, Perez, & Wazer, 2013).

Radiation therapy (or Radiotherapy) plays a central role in dealing with benign and malignant tumors and is an important way to alleviate the clinical symptoms of the disease when treatment is not feasible. Approximately 60% of all cancer patients in Europe and the United States of America are subjected to radiation therapy (external or internal), which either is the main method of treating cancer or used in combination with chemotherapy or surgical removal of the tumor (Halperin et al., 2013). The term radiotherapy usually refers to external irradiation of the tumor region with the aim of delivering the maximum dose to the tumor with the least possible impact on healthy tissue. This can be achieved by splitting and administering the total radiation dose in successive sessions, in order to destroy cancer cells while leaving the possibility of healthy cells to recover.

The conventional radiotherapy treatment includes the following stages: pretreatment imaging, determination of tumor and critical organs, treatment planning, simulation and treatment plan assessment, patient immobilization, quality control and verification, treatment administration. In particular, during the pretreatment imaging, diagnostic images are collected in order to be used to create the treatment plan. Usually, computed tomography (CT) is carried at treatment position for locating the area to be radiated, while other imaging techniques may be used (such as PET and MRI for patients with lung cancer and prostate respectively). The images from CT and possible other imaging modalities are transferred to the treatment planning system, where areas of interest are defined. The areas of interest include the Gross Tumor Volume (GTV), the Clinical Tumor Volume (CTV) and the Planning Tumor Volume (PTV). The GTV is composed of the primary tumor volume, possible metastatic lymphadenopathy or other metastases (GTV M). The CTV comprises the GTV leaving an additional margin (approximately 1cm) for possible microscopic tumor extensions, including all structures to be irradiated. The organ movements can lead to a displacement of CTV, resulting in an additional space (approximately 1cm) that can be added around the CTV to compensate for these uncertainties and the resulting volume is called PTV. Following the determination of areas of interest, treatment planning is performed. During treatment planning, the number and direction of radiation fields (number, irradiation angles, shape, energy, etc.) are determined. The radiotherapy planning system calculates at each pretreatment CT image the dose distribution. Modern treatment planning systems also produce radiographic images of the anatomy, called Digitally Reconstructed Radiographs-DRR. A DRR includes the treatment plan and can be used

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

for accurate simulation and control of geometric data of each field before irradiation. There are two approaches for performing simulation. The first approach involves using a diagnostic x-ray unit (physical simulator) that emulates a treatment machine's (linear accelerator) geometrical movement. The physical simulator provides a real-time image using either a fluoroscopic screen with image intensifier or an amorphous silicon panel. Furthermore, a radiographic film can be produced to provide localization of the treatment volume, visualization of the treatment field and verification of the plan design. The second approach is a combination of the physical process of using a CT scanner as an alternative to the physical simulator and the virtual process of simulator emulation using software. The immobilization is performed by means of dedicated devices which allow for accurate reproduction of patient position from fraction to fraction and ensure comfort and stability. Treatment verification ensures that throughout a treatment course the field placement stays within the intended PTV. This achieved by means of two-dimensional (2D) images acquired using portal film and more recently the use of electronic portal images (EPIs), at megavoltage energies. These images are compared with the reference DRR images in order to detect systematic position errors.

The efficiency of radiotherapy treatment depends mainly on two factors: the adequacy of the radiotherapy plan and the accuracy of the plan implementation in the treatment delivery. An adequate radiotherapy plan involves exact delineation of the target volume as well as precise calculation of the irradiation dose. On the other hand, the efficiency of the treatment delivery pertains to the accuracy with which the radiotherapy plan is reproduced under the treatment conditions. Primary radiotherapy issues, such as target volume coverage and protection of healthy tissue, are closely related to the reproducibility of the plan setup in the different coordinate space of the treatment machine (Aird & Conway, 2002).

This chapter presents the state of the art methods that verify patient position during radiotherapy. Furthermore, an automatic registration method for patient setup verification will be presented. The proposed method performs alignment of the radiotherapy planning CT to an additional verification CT scan, acquired after marking the isocenter position on the patient's skin and before the treatment delivery. The calculated displacement between the isocenter position marked on the patient's skin and the planned isocenter position provides estimation of the required patient position shift before treatment. There are two major advantages of embedding the proposed CT alignment method in the clinical practice of radiotherapy. Firstly, the isocenter displacement between the aligned CT volumes validates the isocenter error correction, estimated by the DRR/EPI comparison. Since the CT volumes contain image information which cannot be represented in the projection images of DRR and EPI, the extent to which the isocenter displacement agrees with the DRR/EPI results, indicates the fidelity of the treatment delivery to the radiotherapy plan. The second advantage is the reduction in time expenditure per patient in the treatment room and improvement of the efficiency of the radiotherapy department, based on the isocenter displacements needed for the patient setup correction, as estimated by the proposed method.

BACKGROUND

In recent years, improvements in imaging technology have impacted on radiotherapy, increasing the accuracy and reliability of the treatment plan and delivery (Aird & Conway, 2002; Driver & Dobbs, 2004; Kirby & Glendinning, 2006). Since Sherouse, et al., (1990) first proposed a computed tomography (CT) - based system for radiotherapy planning, virtual simulation or CT simulation has widely been adopted in radiotherapy clinical practice and has evolved to the standard means of radiotherapy planning,

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

replacing the conventional simulation (Creutzberg, et al., 1992). Virtual simulation enables the accurate delineation of the tumor volume and the radiosensitive normal tissue, such as organs at risk (OAR), and the visualization of the field apertures in relation to the CT dataset. It also provides capabilities of reconstruction of multi-planar images (multi-planar reconstructions, MPR) and determination of specific parameters of the treatment beams, such as field sizes and gantry angles. The previous capabilities are exploited in modern complex radiotherapy techniques, such as conformal radiotherapy (CFRT) and intensity-modulated radiotherapy (IMRT) (Aird & Conway, 2002; Creutzberg, et al., 1992).

Another advantage of virtual simulation is the capability of generating digitally reconstructed radiographs (DRRs) from the planning CT, which, in comparison with electronic portal images, ensure the position of the treatment field and the field coverage of target tissues (Driver & Dobbs, 2004). The comparison of DRR and EPIs constitutes one requisite tool for the accurate "translation" of the planning CT scan localization setup into the treatment position setup of the patient in linear accelerator (LINAC) (Aird & Conway, 2002). The accuracy of this "translation" determines the extent to which the isocenter coordinates of the actual treatment field coincide with the isocenter coordinates defined in the CT volume of the treatment planning. Several studies report the accuracy of the patient setup during treatment, the adequacy of virtual simulation or the need for a conventional simulation following the virtual simulation. The patient setup errors can be easily monitored by portal imaging (Asvestas, Delibasis, Mouravliansky, & Matsopoulos, 2007; Buchali, et al., 2001; Creutzberg, et al., 1996; Creutzberg, Althof, Huizenga, Visser, & Levendag, 1993). In general, these errors depend strongly on the quality of the immobilization techniques and devices (Robar, Clark, Schella, & Kim, 2005). As reported by Schüller et al. (2002), the accuracy of patient repositioning after virtual simulation, calculated by a subsequent conventional simulation, is relatively low in abdominal tumors and breast cancer tumors due to their mobility compared to others such as brain, head or neck tumors, where the accuracy of irradiation can be significantly improved by the usage of immobilization devices (Haddad, et al., 2006; Jiang S. B., 2006). Mostly in thorax and abdomen, tumors are susceptible to artifacts induced by intra-fraction or inter-fraction motion, such as patient motion, organ filling, respiration and cardiac motion. The CT scan for virtual simulation is also affected by the tumor motion, as it is usually performed under the condition of light patient breathing.

The cardinal role of patient setup verification in the treatment procedure has led to numerous studies proposing image processing techniques, which support, facilitate or automate the radiotherapists' work in interpreting the combined information from DRRs and EPIs. Jiang et al. (2008) present an interactive tool for radiotherapists to assist the alignment of portal and simulation images with improved accuracy. Asvestas et al. (2007) present a software-based registration system for setup verification using DRR and portal images, acquired prior to each treatment fraction. Furthermore, methods for the registration of portal images and volumetric CT have been proposed. Khamene et al. (2006) propose an automatic registration method using multiple portal images and the pre-treatment CT volume. DRRs are generated using the CT volume and are compared against portal images acquired right before treatment dose delivery. In order to speed up the registration, the proposed registration method was implemented on graphics processing unit (GPU). Chen et al. (2008) have presented an algorithm for automatic pre-treatment validation. Their algorithm incorporates a hybrid cost function based on partial DRRs and a level set term for similarity measurement and a fast search method by means of parabola fitting and sensitivity-based search order. Fu and Kuduvalli (2008) have proposed an image registration method to perform initial patient setup and frequent detection and correction for patient movement during image-guided cranial radiosurgery treatment. Their method uses DRRs from a preoperative computed tomography volume prior to treatment and x-ray images acquired during patient setup and treatment. Wu et al. (2009) have compared seven

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

similarity measures for use in the 2D-3D registration for patient positioning. The comparison was carried out using both the kilovoltage (kV) and the megavoltage (MV) images of an anthropomorphic cranial phantom and the MV images of a head-and-neck cancer patient. Normalized cross correlation had the best accuracy and robustness among the similarity measures. An extensive review of 2D-3D registration methods for image-guided interventions can be found in (Markelj, Tomaževič, Likar, & Pernuš, 2012). Finally, an approach based on registration of 3D ultrasound images for patient positioning during radiation therapy of prostate has been proposed by (Kaar, et al., 2013).

METHODS

Patient Characteristics and Data Acquisition

For the present study, 40 patients were recruited, treated at the Department of Radiotherapy (RT), Medical School of University of Athens, at Attikon Hospital in Greece, from June until October 2010. The patients' tumors were localized either in the pelvic area (20 patients with prostate cancer) or in the breast/thoracic area (12 patients with breast cancer and 8 patients with lung cancer). The patients with prostate cancer (stage T3N0M0) were referred to the RT department for radical radiotherapy in terms of 74 Gy prescribed dose (2Gy per fraction, 5 days a week). The patients with breast cancer were referred for adjuvant treatment after tumor dissection and chemotherapy (stage T2N0M0). The patients with lung cancer (stage T3N2Mx) underwent radical RT with 60-64Gy prescribed dose (2Gy per fraction, 5 days a week).

For the needs of the present study, two CT scans were acquired from each patient. The first CT scan was required for the radiotherapy planning, whereas the second CT scan was performed prior to the treatment delivery for the verification of the patient positioning in the treatment room. The CT volumes were acquired using a GE HiSpeed Scanner (120.0 kV, 130.0 mA). Slice thickness and slice spacing was 3.0 mm in the abdominal CT scans and 5.0 mm in the breast/thoracic CT scans. Acquisition matrix was 512×512 and the pixel size was approximately 0.889×0.889 mm^2. The number of slices of the radiotherapy planning CT scans ranged from 40 to 100, depending on the tumor position and geometry. The verification CT scan consisted of 11-15 slices with the same slice thickness and spacing as the planning CT scan. Radiotherapy planning was performed with a 3D conformal treatment planning system (VARIAN ECLIPSE ver. 7.1).

For the patient positioning verification before the treatment delivery, orthogonal EPIs and corresponding DRRs were also used. The EPIs were generated from the AGFA CR-25.0 system for electronic portal imaging. For the purpose of image transferring, a workstation of the AGFA system was connected via a Local Area Network (LAN) with the VARIAN ECLIPSE planning system. The DRRs produced from the VARIAN ECLIPSE were transferred and then superimposed on the corresponding EPIs in the AGFA workstation. All the necessary corrections for the patient setup errors using DRR/EPI comparison were made manually.

The Proposed Radiotherapy Procedure

According to the clinical protocol of the study, the patients underwent a CT scan, requisite for the radiotherapy planning. The patients were positioned on the CT scanner couch in the treatment position.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

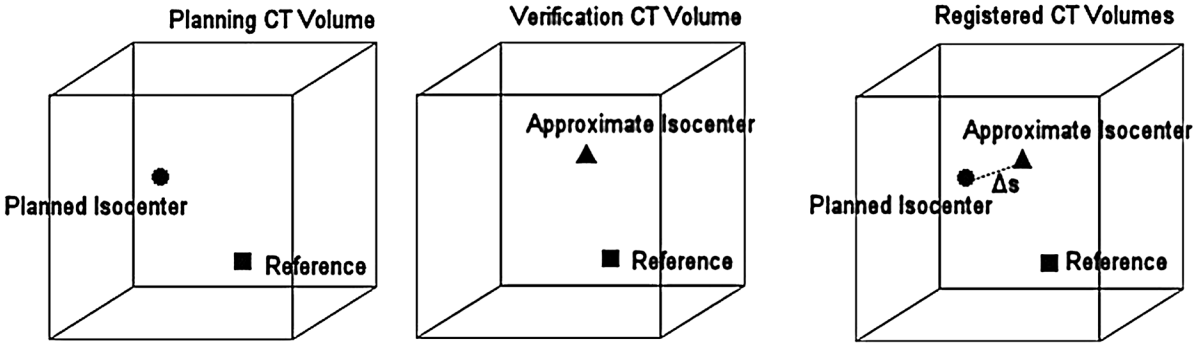
A reference was marked on the patient's skin by the CT operator at the projection points of the laser system of the CT scanner. The acquired CT scans were transferred to a computer workstation (VARIAN ECLIPSE), where contouring and physical treatment planning were performed by the oncologist and the physicist, respectively. The treatment plan was then verified using DRRs in virtual simulation (PROSOMA, π-medical). The RAS (Right-left Anterior-posterior Superior-inferior) coordinates of the planned isocenter were calculated with respect to the reference mark on the patient's skin (Figure 1(a)).

Unlike the standard procedure of radiotherapy, the patients underwent an additional verification CT scan, before the first treatment fraction. This verification CT scan was centered at the isocenter position, as calculated from the planned isocenter RAS coordinates with respect to the reference mark on the patient's skin. The isocenter coordinates in the longitudinal axis of the patient as well as in the transversal plane were also marked by the technologists on the patient's skin by using fiducial radio-opaque markers prior to the CT scan. This marked isocenter of the verification CT volume may be considered as an "approximate" isocenter (Figure 1(b)), since it approximates the planned isocenter position.

After the acquisition of the verification CT, an image registration procedure was employed for the alignment of the two CT volumes. The registration procedure is thoroughly described in the following section. The "approximate" isocenter position in the aligned verification CT scan, marked on the patient's skin, was then compared to the planned isocenter position, defined in the planning CT scan (Figure 1(c)). The displacement of the "approximate" isocenter, marked on the patient's skin, with respect to the planned isocenter, provided an estimation of the positioning errors of the patient's setup for the treatment delivery.

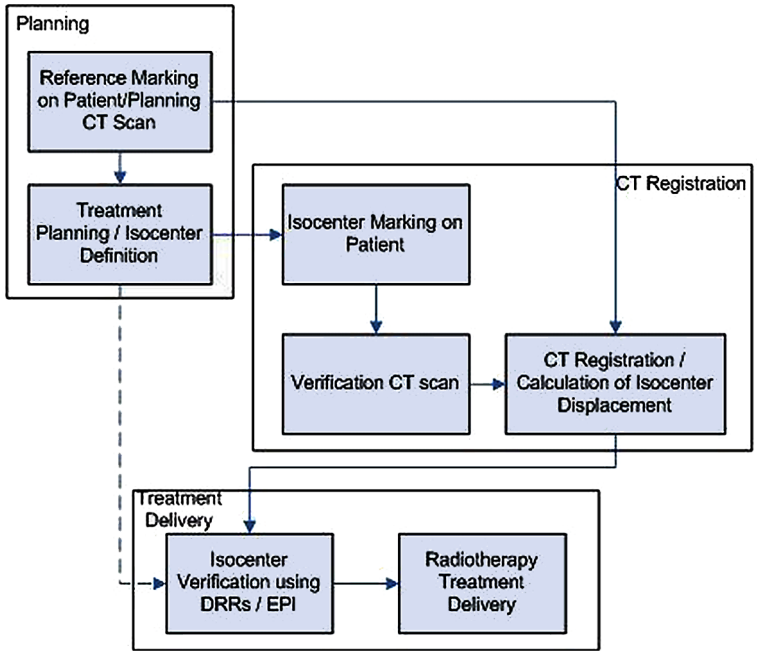
The proposed radiotherapy procedure is represented in the flowchart of Figure 2. As can be seen, the registration of the two CT scans for the verification of the isocenter position is embedded in the radiotherapy procedure as a transitional step between the treatment planning and the treatment delivery. In the treatment room, the patient's setup errors were corrected using the comparison between DRRs and corresponding EPIs, acquired prior to the treatment delivery. Objective of this study was to prove that the isocenter displacement in *x*-, *y*- and *z*-axis, as calculated by the proposed registration method, approximates effectually the actual setup errors in the treatment vault, as estimated by the DRR and EPI comparison.

Figure 1. Representation of (a) the planned isocenter in the planning CT scan, (b) the "approximate" isocenter in the verification CT scan and (c) the planned and "approximate" isocenter in the registered CT scans.



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 2. Flowchart of the proposed radiotherapy procedure



Automatic CT Registration

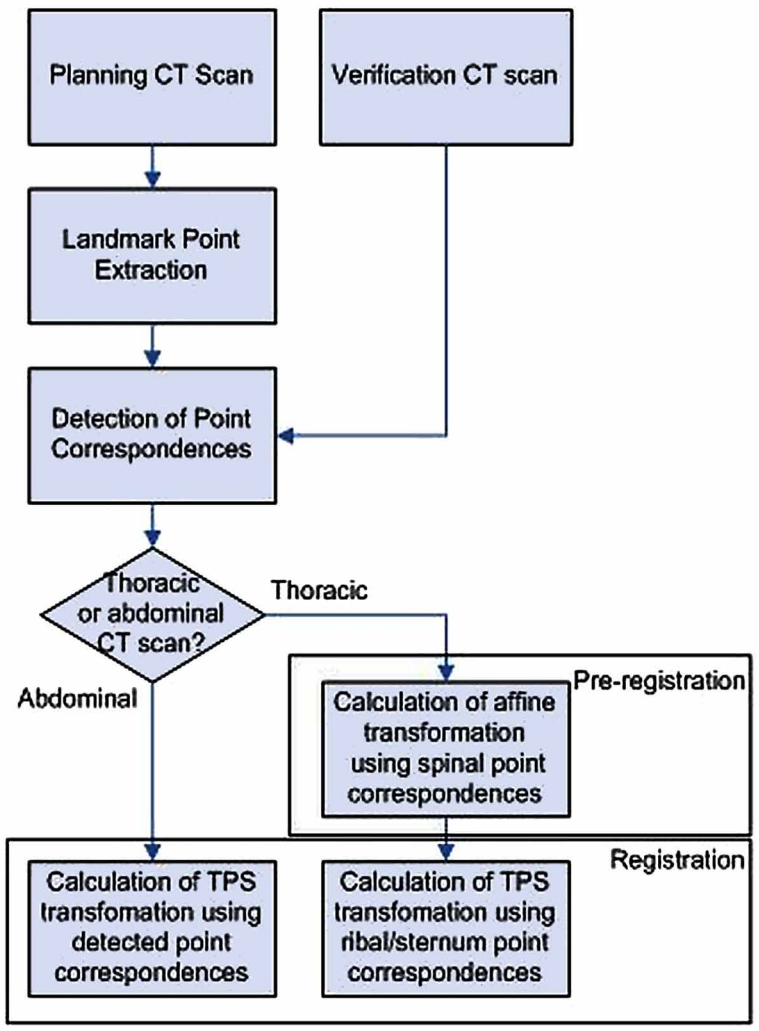
After the acquisition of the two CT scans, an image registration method was employed for the spatial alignment of the CT volumes. The CT scan used for treatment planning was defined as the reference image, whereas the verification CT scan was the image to be registered. The two CT scans were aligned using an automatic point-based registration procedure, the flowchart of which is shown in Figure 3. The registration procedure can be divided into four steps:

- Automatic extraction of landmark points in the planning CT scan.
- Correspondence establishment between homologous points of the two CT scans using an automatic iterative algorithm.
- Pre-registration of the CT scans (required only in the breast/thoracic case) using point correspondences located in the spine.
- Registration of the CT scans using the detected point correspondences.

Prior to the detailed description of the registration procedure, some notations should be introduced. Let I_R and I_F be the reference CT of the radiotherapy planning and the verification CT, respectively. Then, if $\mu_A(I)$ is a restriction of image I in region $A \subset \Re^3$, $T(A) \subset \Re^3$ denotes a local image transformation T acting only on region A of image I.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 3. Flowchart of the CT registration procedure



Automatic Extraction of Landmark Points in the Planning CT Scan

The extraction of landmark points was performed automatically in each CT slice of the reference image I_R of the radiotherapy planning CT scan using the method proposed in (Rohr, 1999). Points of high intensity variations are identified with local maxima of the quantity R, calculated at each pixel (x,y) of the s-slice I_s of the reference image as follows:

$$R(x,y) = \det\left(\overline{\nabla I_s \left(\nabla I_s\right)^T}\right) = \begin{vmatrix} \overline{I_{s_x}^2} & \overline{I_{s_x} I_{s_y}} \\ \overline{I_{s_x} I_{s_y}} & \overline{I_{s_y}^2} \end{vmatrix} = \overline{I_{s_x}^2}\,\overline{I_{s_y}^2} - \left(\overline{I_{s_x} I_{s_y}}\right)^2$$

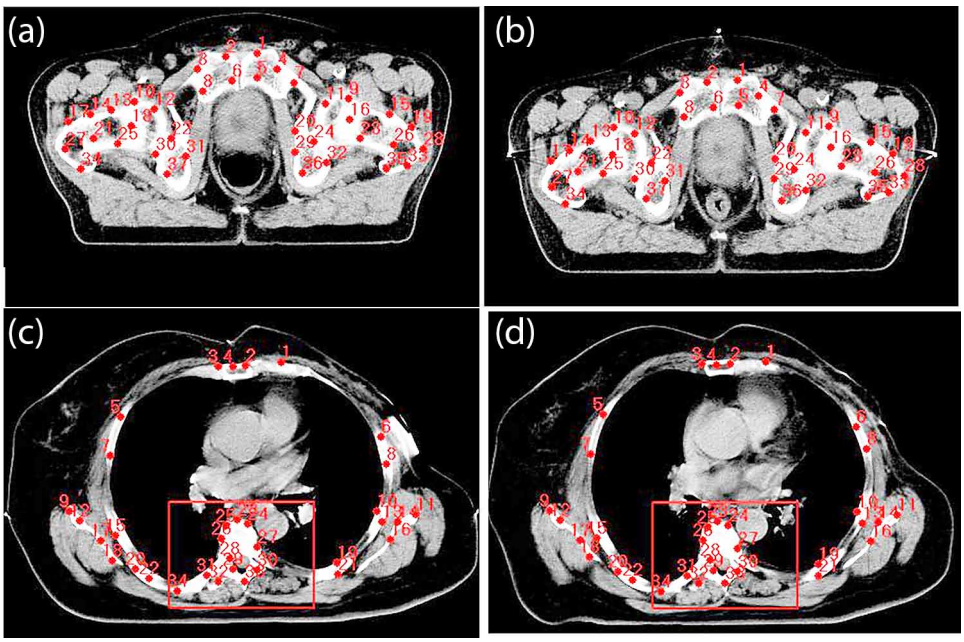
EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

where the overline means average over a local neighborhood, the subscripts denote partial derivatives of the slice I_s with respect to the spatial coordinates x,y, s=1,2,…,M denotes each slice of the reference CT I_R and M is the total number of slices. Since bony structures undergo less deformation than other anatomical features, e.g. soft tissues or organ edges, landmarks located on bones can be more accurately detected in consecutive CT scans. An appropriate threshold, defined after experimentation, was applied to the Hounsfield units of the image at the extracted landmark points, so that only landmark points located on bones were selected. In Figure 4, a typical slice of abdominal (a) and thoracic (c) CT scan is shown along with the numbered landmark points extracted with the proposed method.

Correspondence Establishment between Homologous Points of the CT Scans

For the establishment of point correspondences, an automatic iterative point correspondence algorithm was applied to the two CT volumes. The algorithm was an extension to 3D of the procedure proposed by (Markaki, Asvestas, & Matsopoulos, 2009) for detecting point correspondences in 2D medical images. In particular, local rigid transformations $T_i(A_i)$ were defined, acting only on image blocks A_i centered at the extracted landmark points P_i, i=1,2,…,N, of the reference CT scan I_R. During the iterative procedure, candidate solutions were tested for the parameters of each local transformation T_i, which mapped the region A_i of the reference CT scan I_R to corresponding regions $T_i(A_i)$ of the floating (verification) CT scan I_F. The candidate solutions for the transformation parameters were generated by a random number generator following a "simulated annealing" optimization scheme (Ingber, 1992). Thus, an optimization of

Figure 4. (a) Landmark points in the isocenter slice of an abdominal reference CT scan and (b) their corresponding points in the center slice of the verification CT scan. (c) Landmark points in the isocenter slice of a breast reference CT scan and (d) their corresponding points in the center slice of the verification CT scan. The points located within the rectangular area in (c) and (d) are used for the pre-registration step



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

the similarity between the corresponding image regions A_i and $T_i(A_i)$ in the reference and the verification CT scan was attempted iteratively. The similarity between the restrictions of the two CT volumes I_R and I_F to the corresponding regions A and $T(A)$ was quantified by the correlation coefficient CC as follows:

$$CC\left(\mu_A(I_R),\mu_{T(A)}(I_F)\right)=\frac{\sum_{(x,y,z)\in A}\left[I_R(x,y,z)-\bar{I}_R\right]\left[I_F(x',y',z')-\bar{I}_F\right]}{\sqrt{\sum_{(x,y,z)\in A}\left[I_R(x,y,z)-\bar{I}_R\right]^2\sum_{(x,y,z)\in A}\left[I_F(x',y',z')-\bar{I}_F\right]^2}}$$

where $(x',y',z')\in T(A)$ are the transformed spatial coordinates of $(x,y,z)\in A$ and the overline means average over the specific image region A or $T(A)$.

Figure 4 (b) and (d) show the detected points in verification CTs, denoted with the same number as their corresponding points in reference CTs (Figure 4(a) and (c)) for abdominal and thoracic scans, respectively.

Pre-Registration of the CT Scans (Only in the Breast/Thoracic Case)

The pre-registration of the CT scans was required only in the breast/thoracic case, where the image deformation can be decomposed into two principal components: the deformation due to the rigid movement of the spine and the local deformation due to the respiratory movement. The landmark points located on the spinal vertebrae were considered appropriate for the pre-registration of the CTs so as to determine the parameters of a 3D affine transformation, which compensates the rigid movement of the patient's spine. The affine transformation is defined by twelve parameters:

$$\begin{bmatrix}x'\\y'\\z'\end{bmatrix}=\begin{bmatrix}a_{00}&a_{01}&a_{02}\\a_{10}&a_{11}&a_{12}\\a_{20}&a_{21}&a_{22}\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix}+\begin{bmatrix}a_x\\a_y\\a_z\end{bmatrix}$$

where (x',y',z') are the transformed spatial coordinates of point (x,y,z). The 12 transformation parameters a were calculated using a least squares method (Press, Teukolsky, Vetterling, & Flannery, 1993) between the sets of corresponding points in the two CT volumes (see Figure 4). The pre-registration results can be visually assessed for a typical thoracic CT slice in Figure 5(a)-(b). As can be seen, after the pre-registration step, the edges of the reference image are aligned to these of the transformed image in the region of the spine, indicating an adequate approximation of the rigid movement of patient's body. However, the edges of the two CT scans do not coincide in the front part of the thoracic cavity due to the respiratory movement, which will be compensated in the registration step.

Registration of the CT Scans

The type of transformation used to describe the spatial mapping between the CT scans was a non-rigid transformation, based on the thin plate splines (TPS) model (Bookstein, 1989). The transformation parameters were determined by the corresponding point sets, detected during Step 2 of the procedure. Since pre-registration was not required for the abdominal CT scans, the whole number of corresponding

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

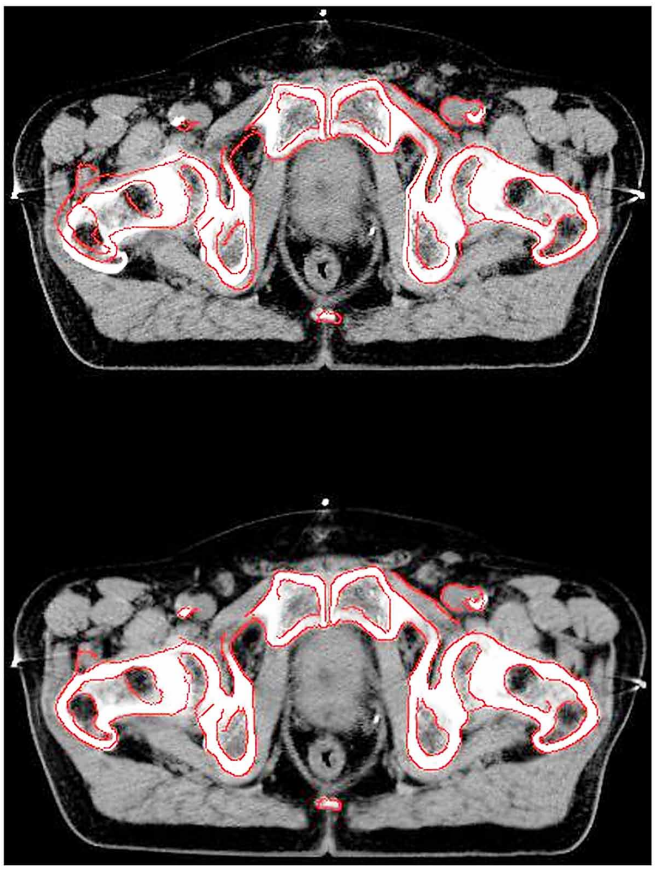
Figure 5. The edges of the reference isocenter slice superimposed on the corresponding CT slice before the registration (a), after the pre-registration (b) and after the registration (c) of the two thoracic CT scans



points was used for the determination of the TPS transformation parameters. For the thoracic CT scans, the points located on the spinal vertebrae were excluded from the registration procedure, since they were used for the determination of the affine transformation of the pre-registration (Step 3). The remaining points, located either on the ribs or on the sternum, were used for the determination of the parameters of the TPS transformation, representing the local deformation, introduced by the respiratory movement of the thoracic walls. In Figure 5(c) and Figure 6(b), the results of the registration are shown for a thoracic and an abdominal CT slice, respectively. The edges of the reference CT slice are superimposed on the aligned CT slice of the verification scan, so that the registration results can be visually assessed.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 6. The edges of the reference isocenter slice superimposed on the corresponding CT slice of the verification CT before (a) and after (b) the registration of the two abdominal CT scans



RESULTS

The isocenter displacements between the CT scan of the radiotherapy planning and the aligned verification CT scan were calculated along the x-axis (lateral), y-axis (anterior-posterior) and z-axis (superior-inferior). The spatial displacement Δs (see Figure 1(c)) of the field isocenter was also calculated using the following equation:

$$\Delta s = \sqrt{(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2}$$

where Δx, Δy, Δz are the displacements along x-, y- and z-axis, respectively. Table 1 and Table 2 contain the results of the isocenter displacement Δx, Δy, Δz calculated by the proposed registration method for the abdominal and the breast/thoracic CT data, respectively.

In order to validate the accuracy of the isocenter displacement results, the patient setup error correction was used, as estimated by the technician from the DRR/EPI comparison prior to the treatment delivery. The DRR/EPI results for the required patient shift in the treatment room are also shown in

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 1. Isocenter displacement calculated by the proposed CT registration method and the DRR/portal comparison along x-, y- and z-axis in abdominal radiotherapy.

Abdominal CT Data	CT Registration			DRR/EPI Comparison		
	Δx (*mm*)	Δy (*mm*)	Δz (*mm*)	Δx (*mm*)	Δy (*mm*)	Δz (*mm*)
1	-0.27	2.86	0.17	-0.30	2.90	0.15
2	-3.71	1.84	-0.44	-3.50	1.70	-0.45
3	2.52	-0.15	0.84	2.50	-0.20	0.85
4	7.64	2.23	0.36	7.50	2.00	0.35
5	-2.64	-6.45	0.70	-2.50	-6.10	0.50
6	3.12	3.81	-0.05	3.00	3.50	0.00
7	2.77	-0.23	1.10	2.70	-0.25	1.00
8	0.00	-4.98	0.42	0.00	-5.00	0.45
9	-5.86	0.73	-0.27	-5.70	0.70	-0.25
10	-3.20	0.96	0.56	-3.20	0.95	0.55
11	2.13	1.87	0.86	2.10	1.80	0.85
12	-4.15	2.56	-0.16	-4.10	2.50	-0.15
13	-0.35	5.47	0.33	-0.40	5.50	0.40
14	-2.83	-1.36	0.20	-2.80	-1.34	0.20
15	-3.81	1.67	-0.29	-3.70	1.65	-0.35
16	-3.17	0.64	0.82	-3.20	0.60	0.80
17	-0.57	-1.48	0.25	-0.60	-1.50	0.20
18	1.84	8.68	0.52	1.80	8.70	0.50
19	0.62	-0.27	-0.08	0.60	-0.25	-0.10
20	5.57	-4.58	-0.05	5.70	-4.70	0.00
Mean	-0.22	0.69	0.29	-0.21	0.66	0.28
St. Dev.	3.53	3.50	0.44	3.47	3.46	0.42

Table 2. Isocenter displacement calculated by the proposed CT registration method and the DRR/portal comparison along x-, y- and z-axis in thoracic radiotherapy.

Thoaracic CT Data	CT Registration			DRR/EPI Comparison		
	Δx (*mm*)	Δy (*mm*)	Δz (*mm*)	Δx (*mm*)	Δy (*mm*)	Δz (*mm*)
1	-8.01	1.06	2.08	-8.00	1.00	2.00
2	1.23	-7.49	0.88	1.20	-7.50	1.00
3	4.56	-0.51	-6.24	4.50	-0.20	-6.00
4	-2.97	-1.78	5.97	-2.95	-1.90	6.00
5	7.05	-6.91	-5.29	7.02	-7.00	-5.50
6	-2.15	-5.43	0.22	-2.20	-5.50	0.00
7	2.54	-5.32	0.67	2.40	-5.10	0.50
8	0.59	2.28	2.70	0.60	2.00	2.50
9	-3.28	-3.22	-2.74	-3.30	-3.00	-2.50
10	-3.63	-7.38	3.56	-3.50	-7.50	3.50
11	-5.96	-7.53	0.84	-6.00	-7.50	1.00
12	-0.31	2.70	2.49	-0.50	2.50	2.50
13	2.03	3.42	-5.23	2.00	3.50	-5.00
14	-1.86	2.34	6.69	-1.80	2.50	7.00
15	-2.78	-8.00	1.61	-2.50	-8.30	1.50
16	-0.25	-4.28	-9.25	-0.50	-4.00	-8.90
17	-1.01	-4.22	1.77	-1.00	-4.50	2.00
18	-5.29	-2.41	5.49	-5.30	-2.50	5.00
19	7.02	-5.56	2.55	7.00	-5.50	3.00
20	2.36	-7.01	5.98	2.50	-7.00	6.00
Mean	-0.51	-3.26	0.74	-0.52	-3.28	0.78
St. Dev.	4.04	3.91	4.40	4.02	3.92	4.33

Table 1 and Table 2 for the abdominal and the breast/thoracic CT data, respectively. The *W*-values of two Wilcoxon signed-rank tests (Corder & Foreman, 2009) for pelvic and breast/thoracic isocenter displacement, respectively, verified that there is no significant difference between the results of the two techniques; hence, the proposed CT registration method provides a good approximation of the isocenter displacement in the treatment position with respect to the treatment plan.

The statistical parameters of the mean values for the displacements Δx, Δy, Δz and Δs were also calculated; the results are presented in Table 3 as well as in Figure 7. As can be seen, the spatial displacement of the isocenter ranged between 0.68 mm and 8.88 mm for the abdominal scans and between 3.59 and 11.20 mm for the breast/thoracic scans. In order to assess these results, a statistical analysis (unpaired T-tests) was performed. Table 3 shows the *p*-values of the T-tests between the abdominal and breast/

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 3. Statistical analysis of the absolute values of isocenter displacement along x-, y- and z-axis in abdominal (a) and thoracic (b) radiotherapy, as calculated by the proposed CT registration method.

	Δ*x* (*mm*)		**Δ*y* (*mm*)**		**Δ*z* (*mm*)**		**Δ*s* (*mm*)**	
	a	**b**	**a**	**b**	**a**	**b**	**a**	**b**
Mean Value	2.84	3.24	2.64	4.44	0.42	3.61	4.46	7.45
Std. Dev.	2.00	2.35	2.33	2.39	0.30	2.49	2.13	2.13
Minimum	0.00	0.25	0.15	0.51	0.05	0.22	0.68	3.59
Maximum	7.64	8.01	8.68	8.00	1.10	9.25	8.88	11.20
Median Value	2.80	2.66	1.86	4.25	0.34	2.72	4.16	7.69
1st Quartile	1.53	1.70	0.90	2.39	0.19	1.73	2.98	5.91
3rd Quartile	3.73	4.74	4.00	6.94	0.59	5.61	5.59	9.05
p-value	0.56		0.02		0.00		0.00	

thoracic isocenter displacements along *x*-, *y*- and *z*-axis as well as the total spatial displacement. The statistical difference between abdominal and breast/thoracic CT scans is insignificant only along *x*-axis (lateral). Since the displacement along *y*- and *z*-axis proves to be significantly higher for breast/thoracic CT scans, it can be inferred that the breast/thoracic radiotherapy is more susceptible to setup errors in these two directions than abdominal radiotherapy, mainly due to respiratory and cardiac motion artifacts.

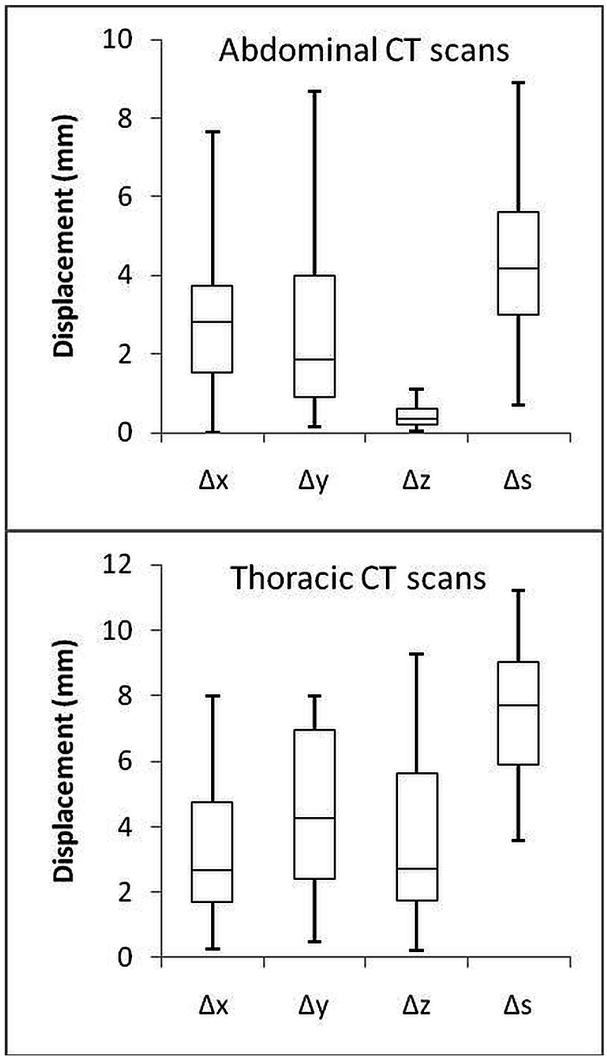
Apart from the verification of the patient positioning errors before treatment, the present study addresses the issue of time sparing in the treatment room, when the proposed CT registration procedure is integrated in the radiotherapy clinical practice. Table 4 summarizes the mean time duration of the radiotherapy procedure, related to the use of either the CT scanner for the treatment planning and the verification or the LINAC for the patient positioning in the treatment room. The total time duration, divided into the CT scanner time and the LINAC time, is given for the standard as well as the proposed radiotherapy procedure. As can be seen, the time required for the CT scanner is increased by 10 min for the proposed procedure; this increase is attributed to the mean time duration of the verification CT scan, which includes the isocenter marking on the patient's skin. However, the time in the LINAC room is significantly reduced in comparison to the standard radiotherapy procedure. This time difference is related to the fact that in the standard procedure the LINAC is used as a typical simulator for the correction of isocenter displacements prior to the treatment delivery, whereas in the proposed procedure its use is limited to the final verification portal just before the initialization of the treatment.

Table 4. Time duration (in min) of the CT simulation and patient position verification.

		Time (Min)	
		Standard Procedure	**Proposed Procedure**
CT	Planning	14	14
	Verification	0	10
LINAC		25	5
Total Time		39	29

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 7. Statistical results of the absolute values of isocenter displacement (in mm) along x-, y- and z-axis in abdominal (a) and thoracic (b) radiotherapy, as calculated by the proposed CT registration method



FUTURE RESEARCH DIRECTIONS

Nowadays, quality assurance in terms of systematic and non-systematic errors of the radiotherapy procedure is one of the main concerns in the field (Kouloulias, 2003). The first objective of this study was to address the issue of radiotherapy quality assurance using an additional verification CT scan, acquired prior to the treatment delivery. The results of the isocenter displacement between the registered verification and planning CT scans were compared with these acquired from the DRR/portal isocenter error correction. This comparison verified the accuracy of the proposed registration method and indicated that it can be used as an auxiliary tool in the radiotherapy clinical practice, in order to provide an approximation of the patient's setup error correction needed before the treatment delivery.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

This automatic setup error correction may prove particularly useful for radiotherapy cases with high repositioning errors, such as abdominal and breast/thoracic radiotherapy, which has been associated in literature with lower reproducibility in patient positioning than head or neck radiotherapy (Schüller, et al., 2002). In abdominal tumors and breast cancer, rates of exact repositioning are found comparatively low; displacements larger than 1 cm can occur between virtual and physical simulation due to several reasons, such as unstable position and obesity with shifting skin marks (Schüller, et al., 2002). These large displacements as well as the non-rigid deformations which are often observed in abdominal and breast/thoracic radiotherapy have led to the adoption of a non-rigid transformation scheme in the proposed method. The non-rigid registration algorithm accounts not only for the rigid motion of the patient between the consecutive scans, but also for soft tissue changes, such as organ deformations and respiratory movements during the treatment.

Regarding the repositioning of the patient in the verification CT scan, it should be noted that it was performed under the same conditions, such as bed, immobilization techniques and devices; thus, the measured displacements were not affected by change of the positioning conditions. As can be seen from the comparative study of the absolute value of isocenter displacement results (Table 3, Figure 7), breast/thoracic CT simulation is more susceptible to positioning errors in all directions (except lateral axis) in relation to the abdominal CT simulation. These errors can be partly attributed to the thoracic movement due to the respiration (Jiang S. B., 2006). Errors in both cases of CT scans occur, due to local deformation of internal tissues or organs (e.g. bladder, bowel). Furthermore, positioning errors prove to be direction-dependent; errors along the longitudinal axis are lower than these along the anterior-posterior or lateral axis.

The second objective was to investigate to which extent the integration of the verification CT scan in the clinical practice of radiotherapy can improve the radiotherapy resource utilization. In the proposed clinical protocol, the procedure of the isocenter marking on the patient's skin takes place in the CT scanner instead of the treatment room. Furthermore, the isocenter displacements, calculated by the automatic registration algorithm prior to the patient's treatment delivery, are indicative of the patient shifts needed for the correction of repositioning errors in the treatment position. The patient setup verification using DRR and EPI comparison is still needed, but the radiotherapy staff has a reliable indication for the necessary shifts of the patient. Thus, if the proposed radiotherapy procedure is adopted, it can contribute to a reduction of approximately 20 min of the time expenditure per patient in the LINAC room (Table 4). This reduction is at the expense of an additional workload of the CT scanner, which lasts about 10 min. However, if the mean time expenditure for the whole procedure of CT simulation and verification is approximately 40 min per patient, the total reduction of 10 min corresponds to a reduction of 25%. In our department the CT scanner is dedicated to radiotherapy and this additional CT workload does not affect the department efficiency. Of course, the staff workload in the CT scanner has also to be taken into consideration.

The European Society of Therapeutic Radiology and Oncology (ESTRO) and the European Institute of Radiotherapy (EIR) have already reported on 3D CT-based in-room image guidance systems in terms of practical and technical guidelines (Korreman, et al., 2010). In our department, we are now establishing a routine procedure for repeating the verification CT once a week in order to further minimize the set-up errors by using the 3D-based information for the isocenter verification, as it is assessed by the proposed method. It is also our feeling that the suggested procedure would potentially be used as an alternative 3D-based image guided technique for departments which do not have in-room CT-based devices.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

CONCLUSION

The proposed procedure provides a reliable radiotherapy validation framework, which can be effectively integrated in the conventional radiotherapy clinical practice. Although modern radiotherapy methods, such us Cone Beam CT-guided radiotherapy, assure radiotherapy accuracy, the accuracy of conventional radiotherapy, which is still used in a large scale worldwide, can be further improved. The fact that the isocenter position can be validated by the proposed procedure in combination with the conventional DRR/EPIs, improves the quality assurance of the whole radiation therapy. The reliability of the radiotherapy procedure is also improved, since the independency between the technologists of the CT-sim and the LINAC eliminates the potential bias of the technologists' subjectivity in case that the radiation therapy was carried out wholly by the same radiotherapy staff. Moreover, it should not be underestimated that the proposed procedure reduces the time expenditure for the LINAC and increases the radiotherapy department efficiency. Beyond that, the method would be potentially used as an alternative procedure for 3D CT-based guided radiotherapy for departments with only 2D-based EPIs mounted in LINACs. Last but not least, in case that the two methods, the proposed registration procedure and the DRR/EPI comparison, provide quite different results of isocenter displacement, systematic errors should be evaluated, analyzed and, consequently, minimized in the whole radiotherapy procedure; this feedback is priceless for the elimination of systematic errors, if any. However, the proposed method needs further validation with more clinical studies.

ACKNOWLEDGMENT

We would like to thank Mr. John Nazos for his kind contribution to the data acquisition.

REFERENCES

Aird, E. G., & Conway, J. (2002). CT simulation for radiotherapy treatment planning. *The British Journal of Radiology*, *75*(900), 937–949. doi:10.1259/bjr.75.900.750937 PMID:12515702

Asvestas, P. A., Delibasis, K. K., Mouravliansky, N. A., & Matsopoulos, G. K. (2007). ESTERR-PRO: A setup verification software system using electronic portal imaging. *International Journal of Biomedical Imaging*, *2007*, 1–10. doi:10.1155/2007/61523 PMID:18521182

Bookstein, F. (1989). Principal wraps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *11*(6), 567–585. doi:10.1109/34.24792

Buchali, A., Geismar, D., Hinkelbein, M., Schlenger, L., Zinner, K., & Budach, V. (2001). Virtual simulation in patients with breast cancer. *Radiotherapy and Oncology*, *59*(3), 267–272. doi:10.1016/S0167-8140(01)00322-X PMID:11369067

Chen, X., Varley, M. R., Shark, L. K., Shentall, G. S., & Kirby, M. C. (2008). A computationally efficient method for automatic registration of orthogonal x-ray images with volumetric CT data. *Physics in Medicine and Biology*, *53*(4), 967–983. doi:10.1088/0031-9155/53/4/010 PMID:18263952

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Corder, G. W., & Foreman, D. I. (2009). *Nonparametric statistics for non-statisticians: A step-by-step approach*. New Jersey: Wiley. doi:10.1002/9781118165881

Creutzberg, C. L., Althof, V. G., de Hoog, M. D., Visser, A. G., Huizenga, H., Wijnmaalen, A., & Levendag, P. C. (1996). A quality control study of the accuracy of patient positioning in irradiation of pelvic fields. *International Journal of Radiation Oncology, Biology, Physics*, *34*(3), 697–708. doi:10.1016/0360-3016(95)02034-9 PMID:8621295

Creutzberg, C. L., Althof, V. G., Huizenga, H., Visser, A. G., & Levendag, P. C. (1993). Quality assurance using portal imaging: The accuracy of patient positioning in irradiation of breast cancer. *International Journal of Radiation Oncology, Biology, Physics*, *25*(3), 529–539. doi:10.1016/0360-3016(93)90077-9 PMID:8436532

Creutzberg, C. L., Visser, A. G., De Porre, P. M., Meerwaldt, J. H., Althof, V. G., & Levendag, P. C. (1992). Accuracy of patient positioning in mantle field irradiation. *Radiotherapy and Oncology*, *23*(4), 257–264. doi:10.1016/S0167-8140(92)80131-2 PMID:1609130

Driver, D. M., & Dobbs, H. J. (2004). Improvements in radiotherapy practice: The impact of new technologies. *Cancer Imaging; the Official Publication of the International Cancer Imaging Society*, *4*(2), 142–150. PMID:18250023

Ferlay, J., Soerjomataram, I., Ervik, M., Dikshit, R., Eser, S., Mathers, C.,... Bray, F. (2013). *GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11*. Retrieved January 30, 2015, from http://globocan.iarc.fr

Fu, D. S., & Kuduvalli, G. (2008). A fast, accurate, and automatic 2D–3D image registration for image-guided cranial radiosurgery. *Medical Physics*, *35*(5), 2180–2194. doi:10.1118/1.2903431 PMID:18561693

Haddad, P., Cheung, F., Pond, G., Easton, D., Cops, F., Bezjak, A., & Wong, R. et al. (2006). Computerized tomographic simulation compared with clinical markup in palliative radiotherapy: A prospective study. *International Journal of Radiation Oncology, Biology, Physics*, *65*(3), 824–829. doi:10.1016/j.ijrobp.2006.01.034 PMID:16682137

Halperin, E. C., Brady, L. W., Perez, C. A., & Wazer, D. E. (2013). *Perez & Brady's principles and practice of radiation oncology* (6th ed.). Philadelphia: Lippincott Williams & Wilkins.

Ingber, L., & Rosen, B. (1992). Genetic algorithms and very fast simulated reannealing: A comparison. *Mathematical and Computer Modelling*, *16*(11), 87–100. doi:10.1016/0895-7177(92)90108-W

Jiang, C. F., Lu, T. C., & Sun, S. P. (2008). Interactive image registration tool for positioning verification in head and neck radiotherapy. *Computers in Biology and Medicine*, *38*(1), 90–100. doi:10.1016/j.compbiomed.2007.07.005 PMID:17825815

Jiang, S. B. (2006). Radiotherapy of mobile tumors. *Seminars in Radiation Oncology*, *16*(4), 239–248. doi:10.1016/j.semradonc.2006.04.007 PMID:17010907

Kaar, M., Figl, M., Hoffmann, R., Birkfellner, W., Stock, M., Georg, D., & Hummel, J. et al. (2013). Automatic patient alignment system using 3D ultrasound. *Medical Physics*, *40*(4), 041714. doi:10.1118/1.4795129 PMID:23556884

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Khamene, A., Bloch, P., Wein, W., Svatos, M., & Sauer, F. (2006). Automatic registration of portal images and volumetric CT for patient repositioning in radiation therapy. *Medical Image Analysis*, *10*(1), 96–112. doi:10.1016/j.media.2005.06.002 PMID:16150629

Kirby, M. C., & Glendinning, A. G. (2006). Developments in electronic portal imaging systems. *The British Journal of Radiology*, *70*(1), S50–S65. doi:10.1259/bjr/21517185 PMID:16980685

Korreman, S., Rasch, C., McNair, H., Verellen, D., Oelfke, U., Maingon, P., & Khoo, V. et al. (2010). The European Society of Therapeutic Radiology and Oncology-European Institute of Radiotherapy (ESTRO-EIR) report on 3D CT-based in-room image guidance systems: A practical and technical review and guide. *Radiotherapy and Oncology*, *94*(2), 129–144. doi:10.1016/j.radonc.2010.01.004 PMID:20153908

Kouloulias, V. E. (2003). Quality assurance in radiotherapy. *European Journal of Cancer*, *39*(4), 415–422. doi:10.1016/S0959-8049(02)00461-6 PMID:12751370

Markaki, V. E., Asvestas, P. A., & Matsopoulos, G. K. (2009). Application of Kohonen network for automatic point correspondence in 2D medical images. *Computers in Biology and Medicine*, *39*(7), 630–645. doi:10.1016/j.compbiomed.2009.04.006 PMID:19481734

Markelj, P., Tomaževič, D., Likar, B., & Pernuš, F. (2012). A review of 3D/2D registration methods for image-guided interventions. *Medical Image Analysis*, *16*(3), 642–661. doi:10.1016/j.media.2010.03.005 PMID:20452269

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1993). Minimization or maximization of functions. In *Numerical Recipes in C: the art of scientific computing* (2nd ed., pp. 412–420). Cambridge University Press.

Robar, J. L., Clark, B. G., Schella, J. W., & Kim, C. S. (2005). Analysis of patient repositioning accuracy in precision radiation therapy using automated image fusion. *Journal of Applied Clinical Medical Physics*, *6*(1), 71–83. doi:10.1120/jacmp.2023.25323 PMID:15770198

Rohr, K. (1999). Extraction of 3D anatomical point landmarks based on invariance principles. *Pattern Recognition*, *32*(1), 3–15. doi:10.1016/S0031-3203(98)00088-0

Schüller, P., Bruns, F., Hesselmann, S., Horn, K., Panke, J. E., Schuck, A., & Willich, N. et al. (2002). Simulator verification of the accuracy of patient repositioning after virtual simulation. Is physical simulation still needed? *Strahlentherapie und Onkologie*, *178*(12), 715–721. doi:10.1007/s00066-002-1037-1 PMID:12491060

Sherouse, G. W., Bourland, J. D., Reynolds, K., McMurry, H. L., Mitchell, T. P., & Chaney, E. L. (1990). Virtual simulation in the clinical setting: Some practical considerations. *International Journal of Radiation Oncology, Biology, Physics*, *19*(4), 1059–1065. doi:10.1016/0360-3016(90)90034-H PMID:2211243

World Health Organization. (2014). *Noncommunicable diseases*. Retrieved from http://www.who.int/topics/noncommunicable_diseases/en/

Wu, J., Kim, M., Peters, J., Chung, H., & Sammant, S. S. (2009). Evaluation of similarity measures for use in the intensity-based rigid 2D-3D registration for patient positioning in radiotherapy. *Medical Physics*, *36*(12), 5391–5403. doi:10.1118/1.3250843 PMID:20095251

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

ADDITIONAL READING

Bankman, I. N. (2000). *Handbook of Medical Imaging*. San Diego: Academic Press.

Brock, K. (2014). *Image Processing in Radiation Therapy (Imaging in Medical Diagnosis and Therapy)*. Boca Raton: CRC Press.

Cherry, P., & Duxbury, A. (2009). *Practical Radiotherapy: Physics and Equipment* (2nd ed.). Oxford: Wiley-Blackwell.

KEY TERMS AND DEFINITIONS

Affine Transformation: A geometric transformation that preserves points, straight lines and planes.

Electronic Portal Imaging: Imaging of radiation field placement using digital means.

Image Registration: A procedure, involving a spatial transformation that brings homologous point in two images into correspondence.

Isocenter: The point in space relative to the treatment machine about which various components of the linear accelerator rotate.

Linear Accelerator: Device used for external beam radiation treatments for patients with cancer.

Radiotherapy: The use of ionizing radiation in order to remove of cancerous tumor or to prevent tumor recurrence after surgery.

Simulator: Medical device used in oncology clinics to perform the simulation for the external beam radiotherapy treatment.

Virtual Simulation: Delineation of the tumor volume and the radiosensitive normal tissue and visualization of the field apertures using CT dataset.

This work was previously published in the Handbook of Research on Trends in the Diagnosis and Treatment of Chronic Conditions edited by Dimitrios I. Fotiadis, pages 211-230, copyright year 2016 by Medical Information Science Reference (an imprint of IGI Global).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 27
Organization and Structure of the Cardiothoracic Intensive Care Unit

Michael H. Wall
University of Minnesota, USA

ABSTRACT

The purpose of this chapter is to emphasize and describe the team nature of critical care medicine in the Cardiothoracic Intensive Care Unit. The chapter will review the importance of various team members and discuss various staffing models (open vs closed, high intensity vs low intensity, etc.) on patient outcomes and cost. The chapter will also examine the roles of nurse practitioners and physician assistants (NP/PAs) in critical care, and will briefly review the growing role of the tele-ICU. Most studies support the concept that a multi-disciplinary ICU team, led by an intensivist, improves patient outcomes and decreases overall cost of care. The role of the tele-ICU and 24 hour in-house intensivist staffing in improving outcomes is controversial, and more research is needed in this area. Finally, a brief discussion of billing for critical care will be discussed.

INTRODUCTION

Modern critical care medicine is the ultimate team sport with all the members of the multidisciplinary critical care team contributing their expertise in the care of the critically ill patient. There is a large body of knowledge and evidence describing and supporting the team nature of critical care delivery.

The objective of this chapter is to review the organization and structure of the CT ICU team emphasizing the importance of the multidisciplinary nature of modern critical care delivery models, physician staffing models, the role of the tele-ICU, the growing field of critical care nurse practitioners and physician assistants (NP/PAs), and finally, will provide a brief overview of critical care billing practices.

DOI: 10.4018/978-1-5225-0549-5.ch027

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

HIGH-RELIABILITY ORGANIZATIONS (ICUs)

Because individual humans and teams can fail, the ideal ICU should be organized and structured in such a way to minimize failure and optimize patient care. Critical care is complex, dangerous, and time sensitive. Other industries such as coal mining, nuclear power, and commercial aviation are also complex, dangerous, and time sensitive, yet in these industries failure cannot be tolerated and the failure rate must be zero. How did these high risk high reliability organizations (HRO) accomplish this and how could this be applied to healthcare? Niedner et al (2013) in a fantastic review article describe how this could be done in a pediatric ICU. Several studies have shown that evidence based care is provided only about 50-60% of the time (McGlynn et al., 2003) and in 2000 the Institute of Medicine's (IOM) "To Err is Human" paper described the ICU (and the operating room) as an area most prone to errors and preventable harm (Kohn, Corrigan, & Donaldson, 2000). To put this in perspective, Larsen et al (2007) showed that in a 20-bed PICU it is estimated that there would be 1416 moderate and 44 serious adverse events per year! Clearly we have room to improve.

Reliability is expressed as the inverse of the failure rate. So if a process fails 1 in 10 times, the failure rate is 10% and the reliability level is 10^{-1}. 10^{-1} systems have very few processes, but rely on individual knowledge and individual work, and training and reminders (like signs in the locker and bathrooms). 10^{-2} systems are a bit better, such as using quality tools and evidence-based protocols (like central line insertion checklists). 10^{-3} systems begin to use even more standardized policies and procedures (Cady, 2008). [See Table 1] Of course the goal would be for all ICUs to perform at the level of six-sigma (3.4 errors in 1,000,000 opportunities), similar to the airline industry. How would we do it? Niedner et al (2013) describe five processes that must all occur to transform an ICU (or anything in healthcare) to a HRO: 1) Preoccupation with failure, 2) Reluctance to simplify interpretations, 3) Sensitivity to operations, 4) Commitment to resilience, and 5) Flexible decision structures.

Table 1. Levels of Reliability

			Opportunities	Real-World	Health Care
Level	**Reliability**	**Success**	**Per Failure**	**Example**	**Example**
Chaotic	$<10^{-1}$	<90%	<10	Annual mortality if >90 years old	Achievement of best-practice processes in outpatient care
1	10^{-1}	90%	10	Mortality of climbing Mt. Everest	Achievement of best-practice processes in inpatient care
2	10^{-2}	99%	100	Mortality of Grand Prix racing	Deaths in risky surgery (American Society of Anesthesiologists grades 3-5)
3	10^{-3}	99.9%	1000	Helicopter crashes	Deaths in general surgery
4	10^{-4}	99.99%	10,000	Mortality of canoeing	Deaths in routine anesthesia
5	10^{-5}	99.999%	100,000	Chartered flight crashes	Deaths from blood transfusions
6	10^{-6}	99.9999%	1,000,000	Commercial airline crashes	—

Note. Levels of Reliability. Adapted from "The High-Reliability Pediatric Intensive Care Unit," by M.F. Niedner, S.E. Muething, and K.M. Sutcliffe, 2013, *Pediatric Clinics of North America, 60,* p. 564. Copyright 2013 by the Elsevier. Adapted with permission.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

In HROs preoccupation with failure means that every small error and near-miss is evaluated and processes to improve are constantly being implemented and evaluated. Sensitivity to operations refers to ongoing attention and real-time data regarding front line operations, such as staffing, supervision, and support of a safety culture. Examples of reluctance to simplify include a culture where diversity in experience and opinion are actively encouraged and every near-miss is evaluated by a multidisciplinary team that is looking for system improvements to these problems. HROs rely on frontline workers' opinions and experience to improve process. Finally, a commitment to resilience means that errors and mistakes will happen, but because of the teamwork and safety culture that exist the team will anticipate failures, overcome them, and learn from them. All ICUs should be organized in such a way to embrace all of these principles of HROs and should all strive to become closer to a six-sigma organization every day (Niedner et al., 2013).

Teamwork and Team Service

The critical care team may consist of intensivists, fellows, residents, medical students, nurse practitioners, physician assistants, nurses, pharmacologists, nutritionists, physical therapists, medical assistants, housekeeping, and others. How does one get this team to work together? Manthous et al (2011) have written an outstanding review article on this topic focusing on four factors that have been shown to impact non-medical team performance that I will briefly review: 1) Leadership style, 2) Psychological safety, 3) Transactional memory, and 4) Mutual accountability. It is interesting how many of these factors are also crucial to making an ICU a six-sigma HRO! Leadership is important for all intensivists, not just medical directors and service chiefs. There are many different leadership styles, but a common model divides into autocratic or transactional and democratic or transformational. It is likely that the situation will determine which style works best. For example, in a rapidly deteriorating patient the autocratic style is probably best. But to develop a HRO where the safety culture is paramount using a democratic or transformational style where the leader seeks out and encourages all members of the team to participate in decision making as needed. Psychological safety is essential in HROs because it allows all members of the team to present different views and question decisions without worrying about being embarrassed in public or fearing retribution. Unfortunately, many physicians exclusively use an autocratic leadership style, which is not effective in enhancing psychological safety. Something as minor as asking each team member to comment on their thoughts for each patient on rounds before the final daily plan is made is one easy way to improve psychological safety. Transactive Memory is a concept that states that the more teams work together and share knowledge and experiences, the better they will function and trust each other as a team. Simulation exercises may be one way to improve transactive memory. Hunziker et al (2009) showed an improvement in CPR performance in teams that had trained together vs teams that had not. Finally, mutual accountability refers to the willingness to accept responsibility for one's actions, and also holding others on the team accountable (Hunziker et al., 2009). A study of top performing academic medical centers found that accountability was key in their success (Keroack et al., 2007). How can these important concepts be implemented in our ICUs? I agree with Manthous et al (2011) who recommended that we should include formal education about teams and leadership for our fellows and residents (and medical students). We should encourage team building exercises along with didactics. They also recommend that formal studies be done to evaluate teamwork and leadership in the performance and outcomes of critical care teams.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

TEAM MEMBERS

There are many members of the team. In this section I will review some aspects of the roles and importance of many members of the team.

Nurse Practitioners and Physician Assistants

Due to the shortage of intensivists and resident work hour restrictions the roles of NP/PAs in the ICU have been expanding. Costa et al (2009) in retrospective cohort database from 2009-2010 and an ICU survey found that 72% of the ICUs reported NP/PAs participating in care of ICU patients. Previous surveys have shown that 24% of PAs and 68% of NPs reported working in ICUs. Costa et al (2014) showed that in 29 mixed medical-surgical ICUs in 22 hospitals where NP/PAs worked in the ICU, the Acute Physiology Scores were lower (42.4 vs 46.7, p<0.001), and the rate of mechanical ventilation was lower (38.8 vs 44.2, p<0.001) when compared to ICUs without NP/PAs. There was no difference in unadjusted and adjusted mortality [adjusted relative risk: 1.10 (95% CI: 0.92-1.31)] which was consistent in all subgroups. This is consistent with other single center studies that showed no difference in outcomes (Costa, Wallace, Barnato, & Kahn, 2014). There are some studies that have shown some improvements in specific outcomes—cost, compliance with bundles, et cetera (Moote, Krsek, Kleinpell, & Todd, 2011). Studies have shown that compliance with evidence based bundles, guidelines, and improved nursing communication (and teamwork) may be responsible for these improvements (Gracias et al., 2008). Several studies have evaluated the role of NP/PAs in trauma ICUs (Collins et al., 2014). These studies have shown that trauma ICU services where NP/PAs were involved had reduced LOS, reduced direct patient costs, improved documentation, and no changes in other quality metrics (D. S. Morris et al., 2012). Collins et al (2014) evaluated the impact of acute care NPs (ACNPs) on a trauma ICU step down service. They found that the addition of ACNPs decreased the hospital LOS from 7.2 to 6.8 days. In addition, there was a $9,000 per patient savings or a $27.8 million dollar per year reduction of charges. Further, a survey showed that 100% of intensivists and staff nurses felt the addition of ACNPs was beneficial and improved flow and communication. Although large multi-center studies are lacking, it appears that the addition of experienced NP/PAs to an ICU service has a very positive effect on morale, communication, and is cost effective. Gershengorn has also written a recent extensive review on this topic (Gershengorn, Johnson, & Factor, 2012).

Critical Care Nursing

Kelly et al (2014) recently showed that ICU nursing staffing, work environment, education, and experience is associated with mortality. Many previous studies on this topic have been inconclusive and have had other methodological problems, which limits conclusions. Kelly et al (2014) evaluated 55,155 older patients being mechanically ventilated that were admitted to 303 hospitals in 4 states (CA, FL, NJ, PA). They then used logistic regression for the effect of ICU nursing variables on 30d mortality after adjusting for patient, hospital, and physician. They found that after adjustment, nursing education with a BS and a better working nurse environment were associated with a decrease in 30d mortality [odds ratios and 95% confidence intervals 0.98 (0.97-0.99), 0.89 (0.81-0.89) respectively]. This translates into ICUs with a better environment had an 11% lower odds ratio for 30d mortality. Further, they found that each 10%

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

increase in the percent of nurses with a BS is associated with a 2% decrease in the odds of 30d mortality. Staffing levels and nurse experience were not associated with mortality. The authors conclude that improving the ICU work environment and hiring more BS trained ICU nurses may improve outcomes. Again, the "work environment" is essential to transition the ICU into a HRO.

Another important part of becoming a HRO is for all leaders to understand what the front line workers actually do. Douglas et al (2013) have reported a detailed evaluation of the unique work nurses perform in the ICU. They performed a total of 147 hours of direct real-time observation in 4 different ICUs at one rural tertiary care community hospital. This study showed that 75% of ICU nurses' time was spent on patient care activities, about 50% of this time was spent on direct patient care (including documentation), 20% on care coordination, 28% on non-patient care, and 2% on indirect patient care. It was also observed that on average nurses performed about 129 activities per hour, averaging a switch between tasks every 29 seconds (Douglas et al., 2013)! What is also surprising is that only 3% of the ICU RNs' time was spent in conversation with ICU doctors and only 2% of their time was spent on bedside rounds with the ICU team. It is no surprise that poor communication with the ICU team is a common problem in many ICUs. I think this type of data is important to know prior to (and after) making changes to the ICU RN workload such as implementation of an electronic health record or new bundles such as oral care, pressure ulcer prevention, et cetera.

Pharmacologists

Preslaski et al (2013) have written an excellent review of the role of the pharmacist as a member of the multidisciplinary ICU team. In my opinion they are an absolutely indispensable part of the team! In 2000 the Society of Critical Care Medicine and American College of Clinical Pharmacy published a position paper describing fundamental, desirable, and optimal activities (Rudis & Brandl, 2000). Fundamental services are those that must be provided to all patients. Desirable activities add things needed for the care of the ICU patient and optimal activities are those needed to optimize outcomes (and required to for the ICU to become an HRO). Unfortunately, two surveys published in 2006, 2007, and 2011 have all showed that fundamental services are being provided <70% of the time, and desirable and optimal services are being provided <1 time per week. This is unfortunate because Bond et al (2007) studied almost 3 million patients across 888 hospitals and showed that hospital mortality decreased as pharmacy staffing improved and pharmacists performed fundamental or desirable services. It is interesting to ponder the effect on mortality if optimal services were provided daily to all patients in every ICU!

The Infectious Diseases Society of America and the Society for Healthcare Epidemiology of America clearly state that pharmacists should be involved in the multidisciplinary antibiotic teams, based on numerous studies showing improved outcomes with pharmacy involvement. For example, pharmacist involvement has decreased length of stay and mortality when they are involved with restricted antibiotic administration, surgical prophylaxis, vancomycin, aminoglycoside dosing, community acquired infections, nosocomial infections, and in sepsis (Preslaski et al., 2013).

Clinical pharmacists have made similar improvements in the management of anticoagulation in the ICU. Again numerous studies have shown that time to achieve target aPTT is shorter, length of stay, mortality, transfusion, number of supratherapeutic INRs and re-admission rates are lower when clinical pharmacists and nursing are independently directing anticoagulation using predefined clinical protocols and pathways (Preslaski et al., 2013).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Similarly studies have shown that pharmacist driven sedation and analgesic protocols have shortened duration of mechanical ventilation, ICU, and hospital LOS. In addition, use of these protocols has increased the time patients were at the goal level of sedation while decreasing drug costs (Preslaski et al., 2013).

Draper et al (2008) showed that more non-compliant AHA-ACLS interventions took place when a pharmacist was not a member of the code team (59.3% vs 31.9% p=0.03). Most of the errors were due to medication errors.

Finally, there is no question that clinical pharmacists decrease the number of adverse drug events (ADEs). Leape et al (1999) reported that the addition of a pharmacist to the ICU team was associated with a decrease of order writing ADE (10.4 vs 3.5 per 1000 patient days, $p<0.001$) and preventable ADEs (33 vs 11.6 per 1000 patient days, $p<0.001$). Other studies have shown that increasing pharmacy staffing from one to five pharmacists per 100 beds reduced ADEs by 48%, and were associated with reduced mortality (Bond & Raehl, 2007).

It is obvious to me that clinical pharmacists improve safety and quality and outcomes in ICU patients. Involvement of ICU clinical pharmacists providing fundamental, desirable and optimal services 7 days a week is an absolutely essential part of the multidisciplinary ICU team for the ICU to become a HRO.

Physical Therapy

Over the last several years there has been growing evidence that supports using less sedation and more physical activity (commonly referred to as "early mobilization") in all ICU patients. Implementation of early mobilization in the ICU requires the expertise of physical (and occupational) therapists, and in the best ICUs they have been added to the multidisciplinary ICU team. Pawlik and Kress (2013) have written an outstanding review on this topic that I will briefly summarize. Several papers have been published describing early mobilization teams that are needed to provide this service. The composition of the teams vary, but to ambulate a patient who is intubated, mechanically ventilated, connected to one or more infusion/nutrition pumps, et cetera takes more than one person! Teams have included: ICU nurses, respiratory therapists, occupational therapists, nursing/rehabilitation assistants, and of course physical therapists. Schweickert et al (2009) in a relatively early study of early mobilization showed that in mechanically ventilated patients who were mobilized early (1.5 days after intubation) vs mobilized late (7.4 days) had improved functional outcomes and increased ventilator free days even though 95% of patients in both groups received mobilization. Morris et al (2008) also showed that early mobilization decreased ICU and hospital LOS. Morris also showed that the cost of the additional members of the mobilization team was about $250,000, but the costs of the intervention group were $500,000 less than the control group. Thus, similar to clinical pharmacist support it appears that the addition of additional members to the ICU team with specialty expertise will improve outcomes—and when looking at the total cost of care for the entre hospitalization may actually save money, even though the individual "siloed" budget for physical therapy (or pharmacy) actually gets worse. Pawlick and Kress (2013) review numerous other studies of early mobilization (and early mobilization teams) that have also shown improvements in ICU team work, culture, multidisciplinary collaboration, and adherence to guidelines such as sedation interruptions. It is interesting to note that similar findings are seen with the addition of clinical pharmacists, and all of these help to improve the "culture" of the ICU, again which is necessary for ICUs to become HROs. Pawklik and Kress (2013) also discuss what is needed to prepare physical therapists to work in the ICU. Studies have shown that physical therapists with >7 years of experience are better suited to work with complex

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

patients in the ICU, but as of 2013 the American Board of Physical Therapy specialties does not offer board certification in acute or critical care physical therapy. Although some hospitals have described competencies for critical care physical therapists, there are no formal training programs, residencies, or fellowships available to train physical therapists with an interest in this growing and important field.

Finally, there is a growing body of evidence that as long as 5 years after recovering from ARDs patients have decreased functional activity, quality of life, and some adverse physiological outcomes, and it is unclear what effect early mobilization and longer term physical therapy after hospital discharge will have on these outcomes. It is known that outpatient rehab programs improve functional capacity and decrease re-admissions after admission for COPD, and improve mortality and quality of life in patients with coronary artery disease and congestive heart failure. It is possible that similar programs following ICU admission may have similar outcomes and more research does need to be done in this area.

Palliative Care

Providing palliative care is a core competency of intensivists. However, extremely complex patients or those not responding to treatment may benefit from a palliative care specialist. Numerous mostly single center studies have shown that early use palliative care services, triggered by ICU specific criterion have increased palliative care involvement, improved patient family and intensivist satisfaction and decreased costs. Nelson et al (2013) recently published a review article describing the selection, implementation and use of screening criterion for palliative care consultation.

Most criterion used triggers focused on family request for consultation, communication problems with the patient or family, and those patients with a high risk of ICU or hospital mortality. Nelson et al (2013) recommend that because each ICU is different, screening criterion should be selected by stakeholders in each ICU and a standard process for implementation of the screening tool, and the impact of the tool should be followed (Nelson et al., 2013). O'Mahony et al (2010) published a study comparing the integration of a palliative care nurse within the ICU team on rounds and with a standardized sign-out to an ICU without such a service. Patients who were in the integrated palliative care model had a formal palliative care consultation happen sooner (2.8 vs 15.5 days, p-0.018), had fewer tests done, and received higher doses of pain medication. They also had more formalized advanced directives and had life prolonging treatments stopped earlier. Most importantly, the addition of the palliative care nurse to the team seemed to improve the multidisciplinary culture of the team and seemed to provide improved communication to the patients.

In summary, similar to clinical pharmacists and physical therapists, the addition of palliative care specialists to the ICU team seems to improve team work, culture, and patient outcomes.

ICU PHYSICIAN STAFFING MODELS

There are literally dozens of articles that have been written on this topic. Garland et al (2013) has written a recent review that summarizes this complex area. In the following sections, I will highlight recent studies that have examined high vs low intensity staffing, 24-hour in-house coverage, staffing ratios and briefly touch on telemedicine.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Closed vs Open ICUs

Closed ICUs are those in which an intensivist is the primary physician directing the care of the patient in the ICU. This model is commonly used in Medical ICUs at academic medical centers in the US and Europe. Open ICUs are those in which multiple specialists care for the patient with or without involvement of an intensivist. Semi-open or semi-closed models also exist where patients are "co-managed" by an intensivist and another physician. This model is quite common in surgical ICUs in academic medical centers in the US and is the model most commonly used in academic cardiothoracic surgical ICUs in the US.

Because the definition of open vs closed does not adequately describe the care model, several studies have compared "high intensity" ICUs where care is transferred to an intensivist or there is mandatory intensivist consultation vs "low intensity" where there is no intensivist involvement. Young et al (2000) and Pronovost et al (2002) both published systematic reviews that concluded high intensity intensivist led care decreased ICU and hospital mortality when compared to low intensity ICUs. However, Levy et al (2008) published a study of 101,000 patients in 123 ICUs using the SCCM's Project Impact data that have showed adjusted hospital mortality was higher in high intensity ICUs (odds ratio 1.42, $p<0.001$). Rubenfeld and Angus (2008) in an editorial entitled "Are Intensivists Safe?" examined the differences between these three systematic reviews. It is interesting that the North American and European guidelines both recommend a high intensity model, ideally with 24-hour in-hospital intensivists (De Lange, Van Aken, & Burchardi, 2002).

Because of these conflicting findings and guideline recommendations, Wilcox et al (2013) performed a systematic review and meta-analyses on fifty-two studies that examined intensivists staffing (high vs low intensity) and ICU or hospital mortality. All 52 trials were observational and involved trials from 13 countries and a variety of ICUs (MICU, PICU, SICU, mixed) from mostly academic medical centers. This study showed that high intensity staffing was associated with decreased ICU and hospital mortality and length of stay. The study also showed that within a high intensity model, 24-hour in-house intensivists did not reduce ICU or hospital mortality. Further this benefit was most beneficial in surgical (risk ratio 0.84; 95% CI 0.44-1.6) and combined medical surgical ICUs (risk ratio 0.76; 95% CI 0.66 – 0.83) vs medical ICUs (risk ration 1.1; 95% CI 0.83-1.5). Furthermore, hospital mortality varied by decade with pooled risk ratios 1980 – 89 (risk ratio 0.74), 1990 – 1999 (risk ratio 0.96), 2000-2009 (risk ratio 0.7) and 2010 – 2012 (risk ratio 1.2).

In an excellent editorial, Capanni et al (2013), pointed out the strengths and weakness of this article. Specifically they suggest that the reason there was no difference seen with 24-hour in-house coverage in high intensity units could be because 49 of the 52 studies were done in academic health centers and there were residents, fellows, and NP/PAs already in the ICU at night. This is an important observation since there is a shortage of intensivists and night work contributes to intensivist burn out.

24-Hour In-House Intensivists

This topic is also controversial because it is recommended by numerous subspecialty societies based on limited observational data, even though the most recent meta-analysis by Wilcox et al (2013) does not support it, and because of the shortage of intensivists, it is almost impossible (and expensive) to provide. Five recent additional studies deserve to be discussed. Gajic et al (2008) showed that the addition of 24-hour in-house intensivists in a high intensity medical ICU (that already had ICUs in-house at night) did

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

not decrease ICU (10.2% vs 10.4% p=0.83) or hospital (17% vs 19% p=0.33) mortality OR/ICU LOS. It also did not improve family satisfaction. It did however increase intensivist burnout.

Garland et al (2012) used an alternating cross-over design in two closed high intensity ICUs (one academic ICU with house staff in house and one community ICU without house staff). Again, there was no improvement in outcomes or family satisfaction, but there were negative effects related to burn out.

Wallace et al (2012) evaluated 49 ICUs that were participating in the APACHE database project. They found that the addition of a nighttime intensivist improves outcomes in low intensity ICUs (odds ratio 0.62, p=0.04) but not in high intensity ICUs (odds ratio 1.08, p=0.78) consistent with Wilcox et al (2013). Kerlin et al (2013) recently reported the first truly randomized prospective trial of 24-hour in-house intensivist coverage in a closed, high intensity MICU at an academic medical center. Intensivists randomly covered seven blocks of nights vs control when they were available by phone. 1,598 patients were studied. The presence of 24-hour in-house intensivists did not decrease ICU LOS (rate ratio 0.98, 95% CI 0.88 – 1.09; P=0.72) or ICU mortality (relative risk 1.07; 95% CI 0.50 – 1.28). It is important to emphasize that this took place in a well-run closed high intensity ICU with residents and fellows in-house at night.

Finally, Kumar et al (2014) evaluated the effect of adding 24-hour in-house intensivist coverage to a high intensity cardiothoracic surgical ICU in patients admitted for more than 48 hours. A previous study by this same group showed that addition of 24-hour coverage reduced blood use, duration of mechanical ventilation and hospital LOS, but did not decrease OR mortality. This study showed similar findings. There was no difference in ICU or 30-day mortality or ICU LOS. Hospital LOS was shorter (12.3 vs 11.0; p<0.01). The 24-hour group also had less transfusions (80.8% vs 65.7%; p<0.001), less sepsis (4.7% vs 0.7%, p<0.01), and less renal failure (22.55 vs 12.5%; p<0.01).

Based on the above studies it is clear that 24-hour coverage does improve many secondary outcomes, but it does not appear to improve mortality. However, most of these studies were under powered to show a mortality benefit, did not look at the effects of 30 day (or longer) mortality or morbidity, and were done in high intensity ICUs. More studies will need to be done to prove 24-hour coverage improves mortality in high intensity models.

Intensivist/Patient Ratios

Ward et al (2013) recently published a statement from the Society of Critical Care Medicine task force on ICU staffing. The task force conducted a review of published literature, held an expert roundtable conference, and an iterative consensus process of the effect of staffing rations (in closed high intensity ICUs) on outcomes. They concluded that the following seven principles should be used when determining an institution's staffing levels: 1) Proper staffing impacts patient care; 2) Large caseloads prolong rounds, limiting time for other work; 3) Staffing decisions need to factor "surge" capacity; 4) Staffing should be routinely reassessed; 5) High staff turnover may be an indicator of excess workload; 6) Other staffing models could be used to decrease the workload of intensivists; 7) In closed academic medical ICUs physician patient staffing levels greater than 1:14 negatively impact education, staff well-being and patient care. This paper also provides a staffing tool that can be used to estimate intensivist staffing needs.

Somewhat related to physician:patient staffing rations is the question of how many intensivists are needed to form a well-functioning ICU group. Reed et al (2014) reported the effects of the number of intensivist FTEs and National Surgical Quality Improvement Program (NSQIP) quality measures. The number of supervised intensivists varied from 8 to 2 over the study period. There was a significant

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

decline in FTEs in 2010. Before 2010 there were an average of 6.6 ± 0.55 vs post decline there was an average of 4.0 ± 1.0 (p=0.0027). Post decline the ICU was often covered by non-intensivists. The period post-decline was associated with an increased risk of mechanical ventilation for > 48 hours (odds ratio 0.94 vs 1.87; p=0.0086) and venous thromboembolism (odds ratio 0.84 vs 1.43; p=0.027). There was no difference in mortality. The quality indicators returned to baseline after full staffing with intensivists was re-established. The authors opine that having an adequate number of intensivists is needed to assure policies, procedures, and protocols are followed. Adherence to these policies improves things that occur outside the walls of the ICU, like venous thromboembolism. They further comment (and I agree) that there must be strong institutional commitment to a fully staffed (surgical in this case) ICU service.

Hand-offs

Due to resident work hour restrictions and some units having 24-hour intensivist models, multiple hand-offs happen every day in all ICUs. There is a large number of studies that have shown associations between hand-offs and medical errors. However, there is another point of view that a "fresh set of eyes" will discover an error (or make a new diagnosis) that was missed. Amaral et al (2014) evaluated the effect of hand-offs by nighttime cross-coverage fellows in an academic mixed medical surgical high intensity closed ICU. In an evaluation of 689 ICU patients admitted for >48 hours they noticed that hand-offs/cross-coverage decreased ICU mortality (odds ratio 0.77; 95% CI 0.60-0.99; p=0.04). Fellows also increased nighttime decision making, transfusions, and pain medicine administration. The authors suggest that a "second look" by fellows may mitigate cognitive errors and actually improve patient safety. In an accompanying editorial, Halpern (2014) comments that although the post-test probability that cross-coverage truly saves lives is quite low—it is certainly possible that "new providers with fresh perspectives in the care of complex patients could yield benefits." Definitely more research on various staffing models and hand-offs need to be done.

Tele-ICU

Another method of providing intensivist coverage is tele-ICU, where a remote ICU team usually consisting of an intensivist, ICU nurses, and some also include software that analyzes data and can act as an early warning system. Hundreds of hospitals in the US have implemented these expensive systems in spite of conflicting evidence supporting their use. Young et al (2011) recently reported a systematic review and meta-analysis on the patient outcomes of telemedicine. They evaluated 13 studies from 35 ICUs. All of the studies were observational and all used before vs after design. The studies included 41,774 patients. Tele-ICU was associated with a decrease in ICU mortality (odds ratio 0.88; 98% CI 0.66-0.77; p=0.02), but not hospital mortality (odds ratio 0.82; 95% CI 0.65-1.03; P=0.08). Tele-ICU was also associated with a decrease in ICU LOS but not hospital LOS. The authors comment that there was huge variation in how tele-ICU was defined, the hospitals it was used in (academic vs community), and the type and staffing model (open vs closed, high vs low intensity) of the ICU it was used in. Based on this study the authors were not able to provide recommendations for which hospitals, ICUSs, or patients are likely to benefit the most from this technology.

Yager et al (2012) evaluated the impact and feasibility of tele-ICU communication in the pediatric ICU. In this study the intensivist was on call from home and was available for telecommunication. 56

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

episodes of communication were evaluated. Telecommunication allowed visual patient assessment, communication with the patient's family, and communication with the entire ICU team, which could not have happened over the telephone alone. Importantly, medical management was changed 32% of the time. The authors commented that tele-communication is feasible, improves team communication, and at least in this study led to changes in management.

Finally Kruklitis et al (2014) have written an excellent review article on the clinical and financial implications for implementation a tele-ICU program. They conclude that the vast majority of research on tele-ICU has shown improvements in mortality, LOS, and cost. In addition, they point out that tele-ICU improves compliance with evidence-based bundles and protocols which are also likely to improve outcome. The authors also perform a very thorough evaluation of the financial impacts of such a system. They conclude that even though installation and yearly operations costs are enormous ($7,000,000 initially and $3,000,000 per year at University of Massachusetts Medical Center) the cost savings to the hospital and payors was $25,000,000 and $13,000,000 per year respectively. This is clearly enough money to pay for the initial capital expense and ongoing operating costs.

BILLING

For an exhaustive review of critical care billing I refer readers to the excellent SCCM monograph edited by Sample and (*Coding and Billing for Critical Care, A Practice Tool*). I will briefly review adult critical care billing in this section. There are two adult critical care codes 99291—for 30-74 minutes, and 99252—for each additional 30 minutes before 74 minutes. 99291 can be used once per day by a group or individual physician for the total time, even if it is not continuous. 99252 is used for time greater than 74 minutes per day. Medicare regulations are clear: providers are not required to be present at the bedside to bill for critical care services. Providers may bill for time spent reviewing patient data, case planning with consultants, writing notes, and discussing patient issues with family members. Providers may bill for their services when they are "immediately available" and in the same unit or floor as the patient. CMS defines critical care as the "direct delivery by physician(s) of medical care for a critically ill or injured patient." This generally must include all three of the collowing criterion: 1) Impairment of one or more organ systems with a high probability of imminent or life-threatening deterioration; 2) Complex decision making to treat or prevent single or multiple organ system failure; 3) Time as described above.

If the initial visit is less than 30 minutes, or the patient is not critically ill then the following evaluation and management codes (E/M) should be used:

99221/level 1 initial hospital care 30 minutes face time
99222/level 2 initial hospital care 50 minutes face time
99223/level 3 initial hospital care 70 minutes face time

There are strict criterion that must be followed to qualify and use these billing codes. See page 73 (*Coding and Billing for Critical Care, A Practice Tool*).

For follow up visits of patients who do not meet critical care codes subsequent hospital visit E/M codes should be used:

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

99231/level 1 subsequent hospital visit 15 minutes face time
99232/level 2 subsequent hospital visit 25 minutes face time
99233/level 3 subsequent hospital visit 35 minutes face time

Additional codes can be used for placement of arterial, central, and PICC lines and for a variety of procedures such as intubation, bronchoscopy, thoracentesis, etc. The reader is referred to *Coding and Billing for Critical Care* for a complete discussion.

FUTURE RESEARCH DIRECTIONS

As I have pointed out in each section much more research needs to be done in all areas in this chapter. Specifically, I think more research needs to be done on how to change the culture of ICUs to become more like HROs. Research needs to be done in how best to train leadership skills during residency and fellowship training.

Much more needs to be done on the longer term outcomes of early mobilization and mobility teams looking at functional outcomes and who benefits the most. Ideally, large studies on the role of 24-hour intensivist models adequately powered to evaluate mortality and costs should be done along with similar studies performed with tele-ICU in both low and high intensity staffing models.

CONCLUSION

Our goal should be to make the ICUs into HROs. To do this, we are going to have to learn and apply leadership and team work skills to build a safety culture that must be present in HROs. I think all ICUs should be staffed in the high intensity staffing model during the daytime. I am not sure that 24 hour staffing is needed in academic medical centers that are staffed 24/7 by fellows, residents, and NP/PAs. The care team model works and has many benefits and I believe NP/PAs should be involved in all ICUs. Further, we should have clinical pharmacologists, physical therapists, and palliative care specialists on all of our ICU teams and these services should be provided 7 days a week. If we do all this we can improve short and long term outcomes for our patients.

REFERENCES

Bond, C. A., & Raehl, C. L. (2007). Clinical pharmacy services, pharmacy staffing, and hospital mortality rates. *Pharmacotherapy*, *27*(4), 481–493. doi:10.1592/phco.27.4.481 PMID:17381374

Cady, R. F. (2008). "Becoming a high reliability organization-operational advice for hospital leaders" report. *JONA'S Healthcare Law, Ethics and Regulation*, *10*, 33. doi:10.1097/01.NHL.0000300780.65358.e0 PMID:18525400

Capanni, F., & Checkley, W. (2013). Differences in hospital mortality by ICU staffing models: You cannot always get what you want, but sometimes you get what you need. *Critical Care Medicine*, *41*(10), 2433–2434. doi:10.1097/CCM.0b013e31829cb0aa PMID:24060770

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Collins, N., Miller, R., Kapu, A., Martin, R., Morton, M., Forrester, M., & Wilkinson, L. et al. (2014). Outcomes of adding acute care nurse practitioners to a Level I trauma service with the goal of decreased length of stay and improved physician and nursing satisfaction. *J Trauma Acute Care Surg*, *76*(2), 353–357. doi:10.1097/TA.0000000000000097 PMID:24398767

Costa, D. K., Wallace, D. J., Barnato, A. E., & Kahn, J. M. (2014). Nurse practitioner/physician assistant staffing and critical care mortality. *Chest*, *146*(6), 1566–1573. doi:10.1378/chest.14-0566 PMID:25167081

De Lange, S., Van Aken, H., & Burchardi, H. (2002). European Society of Intensive Care Medicine statement: Intensive care medicine in Europe--structure, organisation and training guidelines of the Multidisciplinary Joint Committee of Intensive Care Medicine (MJCICM) of the European Union of Medical Specialists (UEMS). *Intensive Care Medicine*, *28*(11), 1505–1511. doi:10.1007/s00134-002-1504-2 PMID:12583367

Douglas, S., Cartmill, R., Brown, R., Hoonakker, P., Slagle, J., Schultz Van Roy, K., & Carayon, P. et al. (2013). The work of adult and pediatric intensive care unit nurses. *Nursing Research*, *62*(1), 50–58. doi:10.1097/NNR.0b013e318270714b PMID:23222843

Draper, H. M., & Eppert, J. A. (2008). Association of pharmacist presence on compliance with advanced cardiac life support guidelines during in-hospital cardiac arrest. *The Annals of Pharmacotherapy*, *42*(4), 469–474. doi:10.1345/aph.1K475 PMID:18349308

Gajic, O., Afessa, B., Hanson, A. C., Krpata, T., Yilmaz, M., Mohamed, S. F., & Wylam, M. E. et al. (2008). Effect of 24-hour mandatory versus on-demand critical care specialist presence on quality of care and family and provider satisfaction in the intensive care unit of a teaching hospital. *Critical Care Medicine*, *36*(1), 36–44. doi:10.1097/01.CCM.0000297887.84347.85 PMID:18007270

Garland, A., & Gershengorn, H. B. (2013). Staffing in ICUs: Physicians and alternative staffing models. *Chest*, *143*(1), 214–221. doi:10.1378/chest.12-1531 PMID:23276844

Garland, A., Roberts, D., & Graff, L. (2012). Twenty-four-hour intensivist presence: A pilot study of effects on intensive care unit patients, families, doctors, and nurses. *American Journal of Respiratory and Critical Care Medicine*, *185*(7), 738–743. doi:10.1164/rccm.201109-1734OC PMID:22246176

Gershengorn, H. B., Johnson, M. P., & Factor, P. (2012). The use of nonphysician providers in adult intensive care units. *American Journal of Respiratory and Critical Care Medicine*, *185*(6), 600–605. doi:10.1164/rccm.201107-1261CP PMID:22135345

Gracias, V. H., Sicoutris, C. P., Stawicki, S. P., Meredith, D. M., Horan, A. D., Gupta, R., & Schwab, C. W. et al. (2008). Critical care nurse practitioners improve compliance with clinical practice guidelines in "semiclosed" surgical intensive care unit. *Journal of Nursing Care Quality*, *23*(4), 338–344. doi:10.1097/01.NCQ.0000323286.56397.8c PMID:18521045

Halpern, S. D. (2014). Cross-coverage in the intensive care unit: More than meets the "i"? *American Journal of Respiratory and Critical Care Medicine*, *189*(11), 1297–1298. doi:10.1164/rccm.201405-0801ED PMID:24881934

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Hunziker, S., Tschan, F., Semmer, N. K., Zobrist, R., Spychiger, M., Breuer, M., & Marsch, S. C. et al. (2009). Hands-on time during cardiopulmonary resuscitation is affected by the process of teambuilding: A prospective randomised simulator-based trial. *BMC Emergency Medicine*, *9*(1), 3. doi:10.1186/1471-227X-9-3 PMID:19216796

Kajdacsy-Balla Amaral, A. C., Barros, B. S., Barros, C. C., Innes, C., Pinto, R., & Rubenfeld, G. D. (2014). Nighttime cross-coverage is associated with decreased intensive care unit mortality. A single-center study. *American Journal of Respiratory and Critical Care Medicine*, *189*(11), 1395–1401. doi:10.1164/rccm.201312-2181OC PMID:24779652

Kelly, D. M., Kutney-Lee, A., McHugh, M. D., Sloane, D. M., & Aiken, L. H. (2014). Impact of critical care nursing on 30-day mortality of mechanically ventilated older adults. *Critical Care Medicine*, *42*(5), 1089–1095. doi:10.1097/CCM.0000000000000127 PMID:24368346

Kerlin, M. P., Small, D. S., Cooney, E., Fuchs, B. D., Bellini, L. M., Mikkelsen, M. E., & Halpern, S. D. et al. (2013). A randomized trial of nighttime physician staffing in an intensive care unit. *The New England Journal of Medicine*, *368*(23), 2201–2209. doi:10.1056/NEJMoa1302854 PMID:23688301

Keroack, M. A., Youngberg, B. J., Cerese, J. L., Krsek, C., Prellwitz, L. W., & Trevelyan, E. W. (2007). Organizational factors associated with high performance in quality and safety in academic medical centers. *Academic Medicine*, *82*(12), 1178–1186. doi:10.1097/ACM.0b013e318159e1ff PMID:18046123

Kohn, L. T., Corrigan, J. M., & Donaldson, M. S. (2000). *To err is human: building a safer health system* (p. 287). Washington, D.C.: National Academy Press.

Kruklitis, R. J., Tracy, J. A., & McCambridge, M. M. (2014). Clinical and financial considerations for implementing an ICU telemedicine program. *Chest*, *145*(6), 1392–1396. doi:10.1378/chest.13-0868 PMID:24889437

Kumar, K., Singal, R., Manji, R. A., Zarychanski, R., Bell, D. D., Freed, D. H., & Arora, R. C. (2014). The benefits of 24/7 in-house intensivist coverage for prolonged-stay cardiac surgery patients. *The Journal of Thoracic and Cardiovascular Surgery*, *148*(1), 290–297.e296. doi:10.1016/j.jtcvs.2014.02.074 PMID:24703332

Larsen, G. Y., Donaldson, A. E., Parker, H. B., & Grant, M. J. (2007). Preventable harm occurring to critically ill children. *Pediatric Critical Care Medicine*, *8*(4), 331–336. doi:10.1097/01.PCC.0000263042.73539.99 PMID:17417126

Leape, L. L., Cullen, D. J., Clapp, M. D., Burdick, E., Demonaco, H. J., Erickson, J. I., & Bates, D. W. (1999). Pharmacist participation on physician rounds and adverse drug events in the intensive care unit. *Journal of the American Medical Association*, *282*(3), 267–270. doi:10.1001/jama.282.3.267 PMID:10422996

Levy, M. M., Rapoport, J., Lemeshow, S., Chalfin, D. B., Phillips, G., & Danis, M. (2008). Association between critical care physician management and patient mortality in the intensive care unit. *Annals of Internal Medicine*, *148*(11), 801–809. doi:10.7326/0003-4819-148-11-200806030-00002 PMID:18519926

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Manthous, C. A., & Hollingshead, A. B. (2011). Team science and critical care. *American Journal of Respiratory and Critical Care Medicine*, *184*(1), 17–25. doi:10.1164/rccm.201101-0185CI PMID:21471081

McGlynn, E. A., Asch, S. M., Adams, J., Keesey, J., Hicks, J., DeCristofaro, A., & Kerr, E. A. (2003). The quality of health care delivered to adults in the United States. *The New England Journal of Medicine*, *348*(26), 2635–2645. doi:10.1056/NEJMsa022615 PMID:12826639

Moote, M., Krsek, C., Kleinpell, R., & Todd, B. (2011). Physician assistant and nurse practitioner utilization in academic medical centers. *American Journal of Medical Quality*, *26*(6), 452–460. doi:10.1177/1062860611402984 PMID:21555487

Morris, D. S., Reilly, P., Rohrbach, J., Telford, G., Kim, P., & Sims, C. A. (2012). The influence of unit-based nurse practitioners on hospital outcomes and readmission rates for patients with trauma. *J Trauma Acute Care Surg*, *73*(2), 474–478. doi:10.1097/TA.0b013e31825882bb PMID:22846959

Morris, P. E., Goad, A., Thompson, C., Taylor, K., Harry, B., Passmore, L., & Haponik, E. et al. (2008). Early intensive care unit mobility therapy in the treatment of acute respiratory failure. *Critical Care Medicine*, *36*(8), 2238–2243. doi:10.1097/CCM.0b013e318180b90e PMID:18596631

Nelson, J. E., Curtis, J. R., Mulkerin, C., Campbell, M., Lustbader, D. R., Mosenthal, A. C., & Weissman, D. E. et al. (2013). Choosing and using screening criteria for palliative care consultation in the ICU: A report from the Improving Palliative Care in the ICU (IPAL-ICU) Advisory Board. *Critical Care Medicine*, *41*(10), 2318–2327. doi:10.1097/CCM.0b013e31828cf12c PMID:23939349

Niedner, M. F., Muething, S. E., & Sutcliffe, K. M. (2013). The high-reliability pediatric intensive care unit. *Pediatric Clinics of North America*, *60*(3), 563–580. doi:10.1016/j.pcl.2013.02.005 PMID:23639655

O'Mahony, S., McHenry, J., Blank, A. E., Snow, D., Eti Karakas, S., Santoro, G., & Kvetan, V. et al. (2010). Preliminary report of the integration of a palliative care team into an intensive care unit. *Palliative Medicine*, *24*(2), 154–165. doi:10.1177/0269216309346540 PMID:19825893

Pawlik, A. J., & Kress, J. P. (2013). Issues affecting the delivery of physical therapy services for individuals with critical illness. *Physical Therapy*, *93*(2), 256–265. doi:10.2522/ptj.20110445 PMID:23162041

Preslaski, C. R., Lat, I., MacLaren, R., & Poston, J. (2013). Pharmacist contributions as members of the multidisciplinary ICU team. *Chest*, *144*(5), 1687–1695. doi:10.1378/chest.12-1615 PMID:24189862

Pronovost, P. J., Angus, D. C., Dorman, T., Robinson, K. A., Dremsizov, T. T., & Young, T. L. (2002). Physician staffing patterns and clinical outcomes in critically ill patients: A systematic review. *Journal of the American Medical Association*, *288*(17), 2151–2162. doi:10.1001/jama.288.17.2151 PMID:12413375

Reed, C. R., Fogel, S. L., Collier, B. R., Bradburn, E. H., Baker, C. C., & Hamill, M. E. (2014). Higher surgical critical care staffing levels are associated with improved National Surgical Quality Improvement Program quality measures. *J Trauma Acute Care Surg, 77*(1), 83-88; discussion 88. doi: 10.1097/TA.0000000000000279

Rubenfeld, G. D., & Angus, D. C. (2008). Are intensivists safe? *Annals of Internal Medicine*, *148*(11), 877–879. doi:10.7326/0003-4819-148-11-200806030-00010 PMID:18519933

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Rudis, M. I., & Brandl, K. M. (2000). Position paper on critical care pharmacy services. Society of Critical Care Medicine and American College of Clinical Pharmacy Task Force on Critical Care Pharmacy Services. *Critical Care Medicine*, *28*(11), 3746–3750. doi:10.1097/00003246-200011000-00037 PMID:11098984

Schweickert, W. D., Pohlman, M. C., Pohlman, A. S., Nigos, C., Pawlik, A. J., Esbrook, C. L., & Kress, J. P. et al. (2009). Early physical and occupational therapy in mechanically ventilated, critically ill patients: A randomised controlled trial. *Lancet*, *373*(9678), 1874–1882. doi:10.1016/S0140-6736(09)60658-9 PMID:19446324

Wallace, D. J., Angus, D. C., Barnato, A. E., Kramer, A. A., & Kahn, J. M. (2012). Nighttime intensivist staffing and mortality among critically ill patients. *The New England Journal of Medicine*, *366*(22), 2093–2101. doi:10.1056/NEJMsa1201918 PMID:22612639

Ward, N. S., Afessa, B., Kleinpell, R., Tisherman, S., Ries, M., Howell, M., & Kahn, J. et al. (2013). Intensivist/patient ratios in closed ICUs: A statement from the Society of Critical Care Medicine Taskforce on ICU Staffing. *Critical Care Medicine*, *41*(2), 638–645. doi:10.1097/CCM.0b013e3182741478 PMID:23263586

Wilcox, M. E., Chong, C. A., Niven, D. J., Rubenfeld, G. D., Rowan, K. M., Wunsch, H., & Fan, E. (2013). Do intensivist staffing patterns influence hospital mortality following ICU admission? A systematic review and meta-analyses. *Critical Care Medicine*, *41*(10), 2253–2274. doi:10.1097/CCM.0b013e318292313a PMID:23921275

Yager, P. H., Cummings, B. M., Whalen, M. J., & Noviski, N. (2012). Nighttime telecommunication between remote staff intensivists and bedside personnel in a pediatric intensive care unit: A retrospective study. *Critical Care Medicine*, *40*(9), 2700–2703. doi:10.1097/CCM.0b013e3182591dab PMID:22732287

Young, L. B., Chan, P. S., Lu, X., Nallamothu, B. K., Sasson, C., & Cram, P. M. (2011). Impact of telemedicine intensive care unit coverage on patient outcomes: A systematic review and meta-analysis. *Archives of Internal Medicine*, *171*(6), 498–506. doi:10.1001/archinternmed.2011.61 PMID:21444842

Young, M. P., & Birkmeyer, J. D. (2000). Potential reduction in mortality rates using an intensivist model to manage intensive care units. *Effective Clinical Practice*, *3*, 284–289. PMID:11151525

ADDITIONAL READING

Afessa, B. (2010). Tele-intensive care unit: The horse out of the barn. *Critical Care Medicine*, *38*(1), 292–293. doi:10.1097/CCM.0b013e3181b9d4dc PMID:20023465

American Medical Association. (2012). *Current Procedural Terminology: CPT 2013*. Chicago, IL: American Medical Association.

American Medical Association. (2012). *ICD-9 CM: 2013 Code Book for Physicians*. Chicago, IL: American Medical Association.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

American Society of Health-System Pharmacists. (2010). ASHP statement on the pharmacist's role in antimicrobial stewardship and infection prevention and control. *American Journal of Health-System Pharmacy*, *67*(7), 575–577. doi:10.2146/sp100001 PMID:20237387

Angus, D. C., Shorr, A. F., White, A., Dremsizov, T. T., Schmitz, R. J., Kelley, M. A., & Critical Care, S. et al. (2006). Critical care delivery in the United States: Distribution of services and compliance with Leapfrog recommendations. *Critical Care Medicine*, *34*(4), 1016–1024. doi:10.1097/01.CCM.0000206105.05626.15 PMID:16505703

Bray, K., Wren, I., Baldwin, A., St Ledger, U., Gibson, V., Goodman, S., & Walsh, D. (2010). Standards for nurse staffing in critical care units determined by: The British Association of Critical Care Nurses, The Critical Care Networks National Nurse Leads, Royal College of Nursing Critical Care and In-flight Forum. *Nursing in Critical Care*, *15*(3), 109–111. doi:10.1111/j.1478-5153.2010.00392.x PMID:20500648

Centers for Medicare and Medicaid Services. 1997 Documentation Guidelines for Evaluation and Management Services. Retrieved December 26, 2014, from http://www.cms.gov/outreach-and-education/medicare-learning-network-mln/mlnedwebguide/downloads/97docguidelines.pdf

Centers for Medicare and Medicaid Services. Critical Care Visits and Neonatal Intensive Care (Codes 99291 – 99292). Retrieved December 26, 2014, from http://www.cms.gov/Regulations-and-Guidance/Guidance/Transmittals/downloads/r1548cp.pdf

Centers for Medicare and Medicaid Services. Medicare Claims Processing Manual [IOM 100-04]. Retrieved December 26, 2014, from http://www.cms.gov/regulationsandguidance/guidance/manuals/internet-only-manuals-ioms-items/cms018912.html

Centers for Medicare and Medicaid Services. Medicare Enrollment Application--Physicians and Non-Physician Practitioners [Form CMS 855i]. Retrieved December 26, 2014, from http://www.cms.gov/Medicare/CMS-Forms/CMS-Forms/downloads/cms855i.pdf

Gawande, A. (2007). The checklist: If something so simple can transform intensive care, what else can it do? *New Yorker (New York, N.Y.)*, 86–101. PMID:18084821

Gawande, A. (2010). *The Checklist Manifesto: How to Get Things Right*. New York, NY: Penguin.

Gosselink, R., Bott, J., Johnson, M., Dean, E., Nava, S., Norrenberg, M., & Vincent, J. L. et al. (2008). Physiotherapy for adult patients with critical illness: Recommendations of the European Respiratory Society and European Society of Intensive Care Medicine Task Force on Physiotherapy for Critically Ill Patients. *Intensive Care Medicine*, *34*(7), 1188–1199. doi:10.1007/s00134-008-1026-7 PMID:18283429

Harris, K. (2006). Critical Care Competency Program Development and Implementation. *Acute Care Perspectives*, *15*, 16–19.

Haupt, M. T., Bekes, C. E., Brilli, R. J., Carl, L. C., Gray, A. W., Jastremski, M. S., & Md, M. H. et al. (2003). Guidelines on critical care services and personnel: Recommendations based on a system of categorization of three levels of care. *Critical Care Medicine*, *31*(11), 2677–2683. doi:10.1097/01.CCM.0000094227.89800.93 PMID:14605541

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Institute for Healthcare Improvement. (2011). *Improvement Leader's Guide to Sustainability and Spread*. Cambridge, MA.

Institute for Healthcare Improvement. (2011). How-to Guide: Sustainability and Spread. Cambridge, MA.

Institute of Medicine (IOM). (2001). *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, D.C. Institute of Medicine (IOM). (2010). The future of nursing: the Institute of Medicine (IOM) issues report. The future of nursing: leading change, advancing health. *The Prairie Rose*, *79*(4), 6.

Kleinpell, R. M. (2014). ICU workforce: Revisiting nurse staffing. *Critical Care Medicine*, *42*(5), 1291–1292. doi:10.1097/CCM.0000000000000202 PMID:24736341

Kleinpell, R. M., Ely, E. W., & Grabenkort, R. (2008). Nurse practitioners and physician assistants in the intensive care unit: An evidence-based review. *Critical Care Medicine*, *36*(10), 2888–2897. doi:10.1097/CCM.0b013e318186ba8c PMID:18766097

Landrigan, C. P., Rothschild, J. M., Cronin, J. W., Kaushal, R., Burdick, E., Katz, J. T., & Czeisler, C. A. et al. (2004). Effect of reducing interns' work hours on serious medical errors in intensive care units. *The New England Journal of Medicine*, *351*(18), 1838–1848. doi:10.1056/NEJMoa041406 PMID:15509817

Lanken, P. N., Terry, P. B., Delisser, H. M., Fahy, B. F., Hansen-Flaschen, J., Heffner, J. E., & Yankaskas, J. R. et al. (2008). An official American Thoracic Society clinical policy statement: Palliative care for patients with respiratory diseases and critical illnesses. *American Journal of Respiratory and Critical Care Medicine*, *177*(8), 912–927. doi:10.1164/rccm.200605-587ST PMID:18390964

Nolan, T., Resar, R., Haraden, C., & Griffin, F. (2004). Improving the Reliatbility of Health Care. *IHI Innovation Series White Paper*.

Pastores, S. M., O'Connor, M. F., Kleinpell, R. M., Napolitano, L., Ward, N., Bailey, H., & Coopersmith, C. M. et al. (2011). The Accreditation Council for Graduate Medical Education resident duty hour new standards: History, changes, and impact on staffing of intensive care units. *Critical Care Medicine*, *39*(11), 2540–2549. doi:10.1097/CCM.0b013e318225776f PMID:21705890

Patterson, K., Grenny, J., McMillan, R., & Switzler, A. (2002). *Crucial Conversations: Tools for Talking when Stakes are High*. New York, NY: Mc-Graw-Hill.

Pronovost, P., Needham, D., Berenholtz, S., Sinopoli, D., Chu, H., Cosgrove, S., & Goeschel, C. et al. (2006). An intervention to decrease catheter-related bloodstream infections in the ICU. *The New England Journal of Medicine*, *355*(26), 2725–2732. doi:10.1056/NEJMoa061115 PMID:17192537

Rhodes, A., Moreno, R. P., Azoulay, E., Capuzzo, M., Chiche, J. D., Eddleston, J., & Valentin, A. et al. (2012). Prospectively defined indicators to improve the safety and quality of care for critically ill patients: A report from the Task Force on Safety and Quality of the European Society of Intensive Care Medicine (ESICM). *Intensive Care Medicine*, *38*(4), 598–605. doi:10.1007/s00134-011-2462-3 PMID:22278594

Selecky, P. A., Eliasson, C. A., Hall, R. I., Schneider, R. F., Varkey, B., & McCaffree, D. R. (2005). Palliative and end-of-life care for patients with cardiopulmonary diseases: American College of Chest Physicians position statement. *Chest*, *128*(5), 3599–3610. doi:10.1378/chest.128.5.3599 PMID:16304319

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Siegal, E. M., Dressler, D. D., Dichter, J. R., Gorman, M. J., & Lipsett, P. A. (2012). Training a hospitalist workforce to address the intensivist shortage in American hospitals: A position paper from the Society of Hospital Medicine and the Society of Critical Care Medicine. *Critical Care Medicine*, *40*(6), 1952–1956. doi:10.1097/CCM.0b013e318258eef7 PMID:22610197

Sutcliffe, K. M. (2011). High reliability organizations (HROs). *Best Practice & Research. Clinical Anaesthesiology*, *25*(2), 133–144. doi:10.1016/j.bpa.2011.03.001 PMID:21550539

The American Association of Critical-Care Nurses. (2005). AACN Standards for Establishing and Sustaining Healthy Work Environments: A Journey to Excellence. *American Journal of Critical Care*, *14*(3), 187–197. PMID:15840893

The Leapfrog Group. ICU Physician Staffing. Retrieved October 19, 2014, from http://www.leapfroggroup.org/media/file/Leapfrog-ICU_Physician_Staffing_Fact_Sheet.pdf

Truog, R. D., Campbell, M. L., Curtis, J. R., Haas, C. E., Luce, J. M., Rubenfeld, G. D., & Kaufman, D. C. et al. (2008). Recommendations for end-of-life care in the intensive care unit: A consensus statement by the American College [corrected] of Critical Care Medicine. *Critical Care Medicine*, *36*(3), 953–963. doi:10.1097/CCM.0B013E3181659096 PMID:18431285

Valentin, A., & Ferdinande, P. (2011). Recommendations on basic requirements for intensive care units: Structural and organizational aspects. *Intensive Care Medicine*, *37*(10), 1575–1587. doi:10.1007/s00134-011-2300-7 PMID:21918847

KEY TERMS AND DEFINITIONS

Closed ICU: ICUs where care is transferred to a specialist, usually but not always an intensivist.

Critical Care Team: This consists of all members of the team which can include: intensivists, medical students, residents, fellows, NP/PAs, nurses, clinical pharmacologists, physical therapists, social workers, unit assistants, housekeepers, respiratory therapists, palliative care specialists.

High Intensity Staffing Model: An ICU that has all care provided by an intensivist or mandatory consultation by an intensivist.

High-Reliability Organization (HRO): High risk and high reliability organizations that are committed to a zero failure rate.

Low Intensity Staffing Model: An ICU where intensivists are only used if specifically consulted.

Open ICU: ICUs where care is provided by many consultants and generalists. Intensivists may or may not be involved in care.

Team Science: The study of successful teams and leadership strategy that could be applied to critical care teams.

Tele-ICU: A staffing model when a team (usually an intensivist and ICU RN) monitors patients usually from a remote location. These teams usually have no direct patient contact, and often are supported by software that analyzes data and provides an "early warning" system.

This work was previously published in Modern Concepts and Practices in Cardiothoracic Critical Care edited by Adam S. Evans, Gregory E. Kerr, Insung Chung, and Robin Varghese, pages 86-104, copyright year 2015 by Medical Information Science Reference (an imprint of IGI Global).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 28
Ethics

Natalia S. Ivascu
Weill Cornell Medical College, USA

Sheida Tabaie
Weill Cornell Medical College, USA

Ellen C. Meltzer
Weill Cornell Medical College, USA

ABSTRACT

In all areas of medicine physicians are confronted with a myriad ethical problems. It is important that intensivists are well versed on ethical issues that commonly arise in the critical care setting. This chapter will serve to provide a review of common topics, including informed consent, decision-making capacity, and surrogate decision-making. It will also highlight special circumstances related to cardiac surgical critical care, including ethical concerns associated with emerging technologies in cardiac care.

INTRODUCTION

Patients undergoing surgery and those that are critically ill are vulnerable and require special consideration. This chapter will provide a brief overview of ethical principles and discuss common ethical issues that arise during the care of cardiac surgical patients. It concludes by exploring contemporary and emerging ethical problems associated with technological advances in life sustaining therapies.

BACKGROUND

Basic Principles of Medical Ethics

Several decades ago, Beauchamp and Childress described a framework to address ethical issues that arise in medical care (Beauchamp, 1979). It consists of four central principles: beneficence, non-maleficence,

DOI: 10.4018/978-1-5225-0549-5.ch028

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

autonomy and justice. While there are limitations to relying solely on these four principles to address the range of issues encountered in the care of the critically ill, they do provide a foundation for contemporary medical ethics.

Beneficence

Beneficence is the desire to act in the best interest of the patient. The core of medical practice is the desire to heal or, at the least, to ease the suffering of a patient. Physicians should consider whether treatments and procedures truly offer the potential for benefit, and counter this against any potential for burden. The modern practice of medicine can make this calculation more complex.

Non-Maleficence

The well-known primary principle of the ethical practice of medicine is the admonition by Hippoctrates "primum non nocere" – "first do no harm." This principle can be especially important when considering highly invasive surgery, such as cardiac surgery. Again, the potential burdens and benefits of the intervention should be fully explored to ensure a reasonable balance,

Autonomy

The principle of autonomy is the right of an individual to govern decisions made about his or her own body. Autonomy is the fundamental groundwork for the informed consent process.

Justice

The principle of justice refers to fairness in the distribution of healthcare resources, as well as respect for the governing laws. Justice is usually invoked when discussing access to treatments that are limited by availability or cost.

Informed Consent

Obtaining consent for a surgery or invasive procedure is frequently focused on the signing of a consent form. However, the informed consent process is far more complex. The written documentation of permission is the result of a frank and open conversation between practioner and patient. The groundwork to the modern standard of informed consent was laid a century ago. The landmark case, Schloendorf vs. The Society of the New York Hospital, is often cited as the turning point in the doctor patient relationship. The patient, Ms. Schloendorf, presented for an examination under anesthesia to investigate a fibroid uterus. The court transcripts note her explicit refusal of surgery before she was placed under anesthesia (Chevernak, 2014). After she was unconscious, the surgeon performed an examination, and upon diagnosing a tumor, proceeded to perform a hysterectomy. The patient sued, and Judge Benjamin Cardozo found in favor of the defendant and stated in his opinion that "…*every human being of adult years and sound mind has a right to determine what shall be done with his body; and a surgeon who performs an operation without his patient's consent commits an assault, for which he is liable in damage, except in cases of emergency, where the patient is unconscious and where it is necessary to operate before*

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

consent can be obtained" (Chevernak, 2014). In the (then) largely paternalistic medical environment, this was a dramatic shift. In addition to elevating the requirements for permission to perform invasive procedures, it marked a move toward honoring patients' autonomy. Many decades after the Schloendorf case, the informed consent standards were again legally challenged by cases in which patients suffered complications that had not been disclosed prior to surgery. Up until this point, the accepted practice was the "professional practice standard," which is what an appropriately experienced physician in the community would tell the patient about the condition, the therapeutic options available, and the generalized risks and benefits of each therapy. The last vestiges of paternalism were cast off with the rejection of the professional practice standard. What emerged was a new requirement to reveal to patients all risks and complications, even those with little probability of occurring.

The modern standard for informed consent includes three key elements: The patient should 1) be given adequate information regarding the nature and purpose of proposed treatments (risks, benefits, and alternatives), 2) be free from coercion, and 3) have medical decision-making capacity.

Decision Making Capacity

The terms "competence" and "capacity" are frequently interchanged; however, in the strictest sense, they are not equivalent. Competence refers to a legal determination of one's ability to enjoy the rights and responsibilities of adulthood. An individual over the age of 18 that is deemed incompetent must fall under the guardianship of another adult and cannot make decisions for himself be it financial, legal, or medical. Capacity more appropriately describes the ability of a person to make a medical decision for himself.

A major distinction separating capacity from competence is that capacity is a temporally and situationally specific designation. The question of capacity encompasses one particular decision at one particular time. The temporal aspect is easiest to understand. Many things may occur in the care of a patient that impairs their decision-making capacity, such as sedation or anesthesia. Certainly a healthy patient arriving from home to the operating room is likely to retain their decision making capacity; however, while under anesthesia the patient is no longer able to make decisions. Other factors affecting capacity may come into play in the intensive care unit, including critical illness, delirium, hypoxemia, metabolic derangement, etc.

There are four main elements of capacity: 1) ability to communicate a choice, 2) ability to understand the relevant information, 3) appreciation of the nature of the condition, and the risk and benefits of proposed treatment, 4) ability to rationally manipulate the given information (Applebaum, 2007). Communication does not have to be done verbally. Every effort should be made to attempt to communicate effectively with intubated or tracheotomized patients. Writing, letter boards, and lip-reading with specialized lip-reading interpreters have been utilized in these circumstances (Meltzer, 2012). The ability to understand information is often equated with cognitive function, but one should exercise great caution with this interpretation. Approximately one-third of learning disabled patients and almost 50% of Alzheimer's patients retain the capacity to make medical decisions (Sessums, 2011). The mini-mental status exam has a poor correlation with capacity determination, except on the extremes of the scored scaling system (Sessums, 2011). When patients come to a decision that differs from medical recommendation, it is critical to elicit the reason behind the patient's decision. In doing so, incorrect perceptions of risks/benefits or side effects can be dispelled. Most importantly, if a patient can rationally explain how he has come to a decision then it must be honored, even if it conflicts with medical advice. When a patient fails to meet these four elements of capacity, an alternative decision maker must be sought.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Surrogate Decision Makers

In the event that a patient loses capacity, the best alternative is an advanced directive outlining what the patient would want in this event. When that is not available, it is beneficial if the patient has previously completed a health care proxy (HCP) form appointing a health care agent (HCA) to speak on his behalf. The requirements for a health care agent or health care power-of-attorney vary from state to state. A HCA has the legal authority to make medical decisions for an incapacitated patient. Ideally, the HCA should be familiar with the patient's values and expressed wishes regarding specific health decisions. When a specific preference is not known, the HCA is asked to use substituted judgment to say what he believes the patient would decide in the given situation. If this, too, is unknown, then the HCA is expected to make a decision that is in the best interest of the patient.

Often, critically ill patients facing life-determining decisions have not completed a HCP designating a HCA. In this event, a surrogate decision maker must be identified. The legal standing for surrogates varies between states. The most recent compilation of state-by-state criteria is in table 1. However, changes have occurred since 2009, such as the Family Health Care Decisions Act that was passed in New York in 2010. This act newly empowered surrogate decision makers to direct the withdrawal or withholding of life-sustaining treatment. In addition, same sex domestic partners were elevated in rank order of priority to be on par with opposite sex spouses (Swidler, 2010).

Witholding and Withdrawing Life Sustaining Therapies

Most deaths in the ICU occur after a decision to withhold or withdraw life-supporting therapies. Withholding life-sustaining therapy refers to situations in which a medical intervention is *not* performed even though death is likely to follow, such as CPR, intubation, or dialysis. Withdrawing life-sustaining therapy refers to *stopping* currently provided care. The decision to limit or remove life-sustaining therapy is generally made when the patient has a minimal chance of recovery or the burden of treatment outweighs the

Table 1. Spectrum of ethical care at the end of life

	Withhold LST	Withdraw LST	Palliative Sedation	Physician Aid in Dying	Euthanasia
Cause of death	Underlying disease	Underlying disease	Underlying disease[a]	Intervention prescribed by physician and used by patient	Intervention used by physician
Intent/goal of intervention	Avoid burdensome intervention	Remove burdensome intervention	Relieve symptoms	Termination fo patient's life	Termination of patient's life
Legal?	Yes[b]	Yes[b]	Yes	Limited by jurisdiction[c]	No
Abrbreviation: LST, life-sustaining treatment.					
[a] Note "double effect."					
[b] Several states limit the power of surrogate decision makers regarding LSTs					
[c] Legal only in states of Oregon, Washington, Montana, and Vermont in the United States					

Table taken from Swetz KM, Mansel JK. Ethical Issues and Palliative Care in the Cardiovascular Intensive Care Unit. *Cardiol Clin* 2013; 31: 657-668.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

potential benefits. It is important to recognize that withholding and withdrawal are morally equivalent actions, and that withdrawal of life-sustaining therapy is not assisted suicide or euthanasia. This is because there is no intention to end life by removing or limiting a technology; rather, the goal is to allow the underlying disease process and the dying process to progress naturally. Without the intent to harm, there is no criminal act. Table 1 outlines the spectrum of ethical care at the end of life (Swetz, 2013).

One example, in the cardiac population, is end-of-life decision-making with cardiac pacemakers and automatic internal debrillators (AICDs). A patient or his surrogate has the authority to request deactivation/discontinuation of cardiac pacemakers and AICDs if continuation of these devices serves to prolong the dying process and/or is no longer consistent with the patient's goals of care (Kramer, 2012). Regardless of the duration or dependency on the device, these devices may legally and ethically be discontinued if this is consistent with the goal of care (Pellegrino, 2005). For many physicians, concerns about futility, often leads to consideration of withholding or withdrawing life-sustaining therapy.

Futility and Futility Disputes

Futility is an amorphous concept fraught with controversy in the public arena. Experts have defined medical futility combining both a quantitative and qualitative approach, as "the unacceptable likelihood of achieving an effect that the patient has the capacity to appreciate as a benefit" (Schneiderman, 2011). Edmund Pellegrino's analysis of withdrawal of life-sustaining therapy incorporates these two components. The quantitative aspect of futility, which lies within the domain of the physician, addresses the effectiveness of the therapy. The qualitative aspect of futility, which lies within the domain of the patient, addresses the subjective benefit of such therapy (Lampert, 2010). Thus, medical futility cannot be decided without the input of both the patient and the physician. Take the example of patient X, a 76-year-old male with non-ischemic cardiomyopathy secondary to chemotherapy for a primary hematologic cancer. He is admitted to the cardiothoracic ICU in decompensated heart failure refractory to medical management. Patient X has failed multiple lines of chemotherapy, but his disease has progressed slowly. Per his oncologist, he could possibly live a year or longer. A left ventricular assist device (LVAD) is offered as a bridge to destination, given that he is not a candidate for heart transplant secondary to his cancer. Patient X refuses. Although a LVAD would provide Patient X with more time, he offers a capacitated refusal, relaying that he finds the burden of living with the LVAD to be overwhelming and disproportionate to the potential benefits. Thus, qualitative futility has been established because, despite the LVAD being a potentially effective therapy, the patient does not feel it provides sufficient benefit.

In this example, withholding life-sustaining therapy as per Patient X's wishes may, or may not, seem ethically ambiguous to you, however it is important to recognize that your colleagues may not always feel comfortable with withholding (or withdrawal) of life-sustaining therapy. In a study from Germany, Jox and colleagues investigated the attitudes of clinicians regarding the definition of futility, communication surrounding futility, and the continuation of life-sustaining therapy despite recognition of futility. The reasons for continuing futile life-sustaining therapy were largely emotional, stemming from clinicians' feelings of guilt, grief, fear of legal consequences, and concerns about the family's reaction (Jox, 2012). It is important to be sensitive to those around you and to support colleagues, in addition to patients and families, when facing difficult decisions.

Situations may arise in which surrogates for critically ill or dying patients wish to continue providing life-sustaining therapies and invasive interventions, even though they offer the patient no benefit and may

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

have prospect for harm. Conflict, often termed a futility dispute, may develop between the clinical team and the surrogate regarding the plan of care. These situations are complicated and can be frustrating for all involved. We suggest trying to anticipate in advance the prospect for a futility dispute to develop so that steps, such as providing frequent clinical updates and ensuring the surrogate is receiving consistent messages about the patient's poor prognosis, can be taken to mitigate the conflict. When conflict does develop, the following might serve to be helpful.

First, ensure that all members of the clinical team, including consultants, are in agreement about the patient's prognosis and what the options are for care. Inconsistent messages can provide a surrogate with false hope and also generate mistrust of the clinical team. Next, ensure that the surrogate has an accurate understanding of the patient's clinical situation. Then, rather than offer discrete options for care, such as "do everything", "do nothing", or "perform CPR" or "issue a do not resuscitate (DNR) order", engage the surrogate in a broader discussion about goals of care. It can be helpful to elicit whether the patient ever expressed any preferences about their medical care or what they might want, given the current clinical situation. Helping to frame this as a discussion of what *the patient* would want can help to alleviate the surrogate of the burden of medical decision-making. Finally, once the surrogate has articulated a goal of care, i.e. a goal of comfort or a goal to reduce suffering, then craft a plan in conjunction with the surrogate that will serve to achieve this goal. For example, if the goal is to provide comfort, then explain that medications will be provided for palliation and that CPR would be inconsistent with this objective. Depending on your institution, medical ethics, palliative care, and/or pastoral care may be consulted to assist you with this process.

Specific Religious Issues Regarding Withdrawl of Life-Sustaining Therapy

It is often difficult to initiate a discussion of goals of care and end-of-life care with a patient or surrogate. Some understanding of the patient's religious background can aid in providing a framework for this discussion. It is important to note, however, that while a patient may identify himself as a member of a particular religious community, he may not ascribe to all beliefs set forth by that religion. As such, one should proceed cautiously and without presumption when addressing withdrawal of life-sustaining therapies. Below is a brief comment regarding the viewpoints of several, major religions to familiarize the reader. This is meant only as a brief introduction for the critical care practitioner, and not as a comprehensive religious review. It is advised that all patients and surrogates are asked whether they follow a particular religious tradition and whether a member of their own clergy or hospital pastoral care might be a source of support. In addition, local ethics consultation service may also provide assistance.

Christian: Catholic/Orthodox Christian

The Catholic and Orthodox Christian perspectives begin with the principle that the intentional, direct, or voluntary killing of another human being is always wrong. To that end, removing life-sustaining therapies with the intension of ending life is also impermissible and morally equivalent to euthanasia or suicide. However, removing such treatments in the event that they are futile, or that the burden of treatment outweighs the potential benefit, is ethically permissible as a resignation of an inevitable death. One careful distinction in these faiths is that the burden of life is not the same thing as the burden of treatment. As such, if a particular treatment is painful and shows little potential benefit, a Catholic or

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Orthodox Christian can in good conscience forgo that care. However, the intolerability of a difficult or limited life is not justification to terminate life-sustaining measures (Donovan, 1997).

Christian: Evangelical Protestant

There are many other denominations within the Christian faith and this description may not be all encompassing. In general, Evangelical Protestants accept that death is an inevitable end to a human's existence on Earth. They generally hold that all medical options should be exhausted in the defense of life. Once the medical personnel believe that battle has been lost, however, the faith would not object to allowing death to occur. They accept withdrawal of treatment to sustain life with the intention to allow death to occur. It is important for treating clinicians to keep in mind the strongly held belief that judgment in the eyes of God will come with their death. Family members may be conflicted about resigning to death if they fear that forgiveness and peace with God has not yet been obtained (Ankeny, 2005).

Orthodox Jewish

Orthodox Judaism views life as precious and sacred to the end. The defense of life is intended to trump any other mandate. Jewish law, or "Halacha", implores physicians to do anything possible to save a patient, no matter how briefly the patient may survive. Food, fluids, and oxygen may never be withheld even if artificial means are necessary. When death approaches and is clearly inevitable, the person is considered to be "a Goses". A "Goses" is a person actively in the process of dying whose soul has not yet departed. Any action to hasten the death of a Gosses is akin to murder, however palliation and provisions for comfort while dying are allowed, even if they accelerate the process. For example, administration of pain medication is permissible even if respirations are depressed, as long as the intention is only to palliate and not to hasten death (Brenner, 2002; Baeke, 2005; Steinberg, 2003).

Islam

Muslims derive their moral directives from four sources in order of importance: the Koran, the Sunnah, the consensus of the learned, and wisdom. Life is seen as a gift from God, and the central Islamic prophet has been quoted as saying, "seek treatment, for every illness God created a treatment." As such, if there is any hope for recovery, discontinuation of treatment is not permissible. Islamic law forbids any form of euthanasia. A strict interpretation of the law would not allow for withdrawal of life-support, owing to the impossibility of knowing with certainty the outcome of treatment. A modern interpretation does advocate for avoiding prolongation of suffering when there is no hope of survival. However, withholding food and hydration is permitted as this is seen as distinct from the removal of artificial life support (Alsolamy, 2014).

Hindu

There is no central authority or text to dictate the beliefs of the Hindu faith. Traditionally, Hinduism identifies suicide as morally intolerable. Yet modern Hindu practice is generally supportive of stopping life-support in futile situations. There may be a wide variety of views in common practice.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Buddhist

Although suicide and murder are prohibited in Buddhist traditions, ending one's life to help others is a practice ascribed to even Buddha himself. As believers in reincarnation, the impact of death is received differently than that in other faiths. Of primary importance is to die in a calm, conscious state, as this is thought to provide for the most favorable rebirth. There is no clear consensus on Buddhist approach to ending life-support, but in general they do not advocate for the prolonging of life or the hastening of death (Chaicharoen, 1998). A unique feature of Buddhism is the belief that the mind is not the same as the brain. Thus, while the heart remains beating, Buddhists may not easily accept brain death as the death of the person. Buddhists would generally accept palliative care, but with the aim of preserving consciousness while making the patient comfortable (Chan, 2012).

FUTURE RESEARCH DIRECTIONS

Emerging Technologies

Advances in mechanical circulatory support (MCS) and increasing use of extracorporeal life support (ECSL) and left ventricular assist devices (LVAD) have given rise new arenas for ethical debate. While these therapies do have the potential to save the lives of patients who otherwise would certainly die from cardiac failure, they are invasive and burdensome and require careful consideration, if not before initiation of therapy, then certainly after the patient is stabilized (Meltzer, 2014).

Emergency ECLS

Failed cardiopulmonary resuscitation (CPR) has generally been synonymous with certain death. Many cardiac surgery centers, however, now have the capability to rapidly place a patient with cardiac failure on ECLS and maintain oxygenation and perfusion in spite of a failing heart. Despite being a highly invasive procedure, in these circumstances ECLS is often initiated without explicit informed consent by invoking the emergency presumption (i.e. the presumption that a patient would want to receive potentially life saving therapies unless he or his surrogate specifically noted otherwise).. The urgency, which is the very essence of an emergency intervention, prohibits a long and detailed conversation with the patient's surrogate and family. The intention of emergency ECLS is to provide a "bridge to decision." Maintenance on ECLS can prevent secondary organ failure while allowing time to assess neurologic recovery. ECLS may even allow for an opportunity for the patient to regain capacity to make decisions for themselves regarding advanced therapies, such as long-term assist devices or heart transplantation. Advance directives rarely, if ever, address such a specific situation and generally offer little guidance in the emergency setting (Shah, 2014).

Termination of LVAD Support

In spite of technologic advances, heart failure is a progressive and terminal condition. ECLS has been very successful at improving both quality and duration of life in this setting. However, there may come a time at which the patient no longer wishes to continue LVAD support. Depending on the patient's

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

underlying function, termination of LVAD support has a variable outcome, ranging from sudden death to pulmonary congestion with gradual deterioration. The decision to stop a LVAD should be made in concert with a palliative care plan to ensure the comfort and dignity of the patient. The LVAD is a supportive therapy, not a replacement therapy. As such, it is ethically permissible to withdraw such support when the patient or, if necessary, the patient's proxy determines that ongoing treatment would be futile. As with any withdrawal of life-sustaining therapy, local institutional guidelines should be consulted prior to discontinuation.

CONCLUSION

Ethical issues commonly arise in the intensive care unit, particularly concerning including informed consent, decision-making capacity, and surrogate decision-making. Cardiac surgery patients may be incapacitated and unable to make decisions for themselves. They may be ill for prolonged periods of time and questions regarding withholding and withdrawal of life-sustaining therapy may arise. Understanding the legal as well as ethical responsibilities regarding decision-making for this vulnerable population is a cornerstone of critical care. Ethics consultation may serve to assist you in clarifying goals of care and mediating futility disputes.

REFERENCES

Alsolamy, S. (2014). Islamic views on artificial nutrition and hydration in terminally ill patients. *Bioethics*, *28*(2), 96–99. doi:10.1111/j.1467-8519.2012.01996.x PMID:22845721

Ankeny, R. A., Clifford, R., Jordens, C. F. C., Kerridge, I. H., & Benson, R. (2005). Religious perspectives on withdrawal of treatment from patients with multiple organ failure. *The Medical Journal of Australia*, *183*, 616–621. PMID:16336148

Applebaum, P. S. (2007). Assessment of Patients' Competence to Consent to Treatment. *The New England Journal of Medicine*, *357*(18), 1834–1840. doi:10.1056/NEJMcp074045 PMID:17978292

Baeke, G., Wils, J. P., & Broeckaert, B. (2005). 'There is a time to be born and a time to die' (Ecclesiastes 3:2a): Jewish perspectives on euthanasia. *Journal of Religion and Health*, *50*(4), 778–795. doi:10.1007/s10943-011-9465-9 PMID:21253848

Beachamp, T. L., & Childress, J. F. (1979). *Principles of Biomedical Ethics*. New York, NY: Oxford University Press, Incorporated.

Brenner, D. S. (2002). *Embracing life & facing death: a Jewish guide to palliative care*. New York, NY: CLAL.

Chaicharoen, P., & Ratanakul, P. (1998). Letting-Go or Killing: Thai Buddhist Perspectives on Euthanasia. *Journal of Asian and International Bioethics*, *8*, 37–40.

Chan, R. W., & Hegney, D. (2012). Buddhism and medical futility. *Bioethical Inquiry*, *9*(4), 433–438. doi:10.1007/s11673-012-9392-9 PMID:23188402

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chevernak, J., McCullough, L. B., & Chevernak, F. A. (2014). Surgery Without Consent or Miscommunication? A New Look at a Landmark Legal Case. *American Journal of Obstetrics and Gynecology*, *212*(5), 586–590. doi:10.1016/j.ajog.2014.06.062 PMID:24995772

Donovan, G. K. (1997). Decisions at the end of life: Catholic tradition. *Christ.*, *3*(3), 188–203. doi:10.1093/cb/3.3.188 PMID:11655313

Jox, R. J., Schaider, A., Marckmann, G., & Borasio, G. D. (2012). Medical futility at the end of life: The perspectives of intensive care and palliative care clinicians. *Journal of Medical Ethics*, *38*(9), 540–545. doi:10.1136/medethics-2011-100479 PMID:22562948

Kramer, D. B., Mitchell, S. L., & Brock, D. W. (2012). Deactivation of Pacemakers and Implantable Cardioverter-Defibrillators. *Progress in Cardiovascular Diseases*, *55*(3), 290–299. doi:10.1016/j.pcad.2012.09.003 PMID:23217433

Lampert, R., Hayes, D. L., Annas, G. J., Farley, M. A., Goldstein, N. E., Hamilton, R. M., & Zellner, R. et al. (2010). HRS Expert Consensus Statement on the Management of Cardiovascular Implantable Electronic Devices (CIEDs) in patients nearing end of life or requesting withdrawal of therapy. *Heart Rhythm*, *7*(7), 1008–1026. doi:10.1016/j.hrthm.2010.04.033 PMID:20471915

Meltzer, E. C., Gallagher, J. J., Suppes, A., & Fins, J. J. (2012). Lip-reading and the ventilated patient. *Critical Care Medicine*, *40*(5), 1529–1531. doi:10.1097/CCM.0b013e318241e56c PMID:22430239

Meltzer, E. C., Ivascu, N. S., Acres, C. A., Stark, M., Kirkpatrick, J. N., Paul, S., & Fins, J. J. et al. (2014). Extracorporeal membrane oxygenation in adults: A grief review and ethical prconsiderations for non-specialist health providers and hospitalists. *Journal of Hospital Medicine*, *9*(12), 808–813. doi:10.1002/jhm.2262 PMID:25332093

Pellegrino, E. D. (2005). Futility in medical decisions: The word and the concept. *HEC Forum*, *17*(4), 308–318. doi:10.1007/s10730-005-5156-9 PMID:16637443

Schneiderman, L. J. (2011). Defining medical futility and improving medical care. *Journal of Bioethical Inquiry*, *8*(2), 123–131. doi:10.1007/s11673-011-9293-3 PMID:21765643

Sessums, L. L., Zembruzuska, H., & Jackson, J. L. (2011). Does this patient have decision-making capacity? *Journal of the American Medical Association*, *306*(4), 420–427. doi:10.1001/jama.2011.1023 PMID:21791691

Shah, K. B., Levenson, J. L., & Mehra, M. R. (2014). Emergent use of mechanical circulatory support devices: Ethical dilemmas. *Current Opinion in Cardiology*, *29*(3), 281–284. doi:10.1097/HCO.0000000000000058 PMID:24686401

Steinberg, A. (2003). *Encyclopedia of Jewish medical ethics: a compilation of Jewish medical law on all topics of medical interest*. Jerusalem, New York: Feldheim Publishers.

Swetz, K. M., & Mansel, J. K. (2013). Ethical Issues and Palliative Care in the Cardiovascular Intensive Care Unit. *Cardiology Clinics*, *31*(4), 657–668. doi:10.1016/j.ccl.2013.07.013 PMID:24188227

Swidler, R. N. (2010). The Family Health Care Decisions Act: A Summary of Key Provisions. *Health Law Journal*, *15*(1), 32–35.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

KEY TERMS AND DEFINITIONS

Autonomy: The right of persons to make decisions about their fate.

Beneficence: The desire to act in the patient's best interest.

Health Care Agent: A person legal appointed by a patient to make medical decisions should the patient lose capacity.

Heath Care Proxy: A form that a patient completes to legally empowers another individual, ie the health care agent, to make medical decisions should the patient lose capacity.

Justice: The principle of fairness in distribution of resources.

Non-Maleficence: The desire to avoid harm.

Surrogate: A person selected to serve as a decision maker for an incapacitated patient when a health care agent has not been previously appointed. The surrogate is determined by proximity as next-of-kin.

Withdrawal of Life-Sustaining Therapy: The removal of medications or machines, without which a patient is not expected to survive.

Withholding Life-Sustaining Therapy: The decision not to provide a medical therapy, without which the patient is not expected to survive.

This work was previously published in Modern Concepts and Practices in Cardiothoracic Critical Care edited by Adam S. Evans, Gregory E. Kerr, Insung Chung, and Robin Varghese, pages 124-134, copyright year 2015 by Medical Information Science Reference (an imprint of IGI Global).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 29
Risk, Activism, and Empowerment:
Women's Breast Cancer in Venezuela

Mahmoud Eid
University of Ottawa, Canada

Isaac Nahon-Serfaty
University of Ottawa, Canada

ABSTRACT

The prevalence of breast cancer in Venezuela is particularly alarming, which is attributed to healthcare inequalities, low health literacy, and lagging compliance with prevention methods (i.e., screening and mammography). While the right to health is acknowledged by the Venezuelan constitution, activism beyond governmental confines is required to increase women's breast cancer awareness and decrease mortality rates. Through the development of social support and strategic communicative methods enacted by healthcare providers, it may be possible to empower women with the tools necessary for breast cancer prevention. This paper discusses issues surrounding women's breast cancer, such as awareness of the disease and its risks, self-advocacy, and the roles of activists, healthcare providers, and society. Specifically, it describes a four-year action-oriented research project developed in Venezuela, which was a collaborative work among researchers, practitioners, NGOs, patients, journalists, and policymakers. The outcomes include higher levels of awareness and interest among community members and organizations to learn and seek more information about women's breast cancer, better understandings of the communicated messages, more media coverage and medical consultations, increasing positive patient treatments, expansion of networking of NGOs, as well as a widely supported declaration for a national response against breast cancer in Venezuela.

DOI: 10.4018/978-1-5225-0549-5.ch029

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

INTRODUCTION

On May 25th, 2005, the World Health Organization (WHO) recognized cancer as a worldwide public health problem; specifically, breast cancer existing as critical (Cazap et al., 2010). Due to the global impact of this disease, research surrounding its prevalence and impact has become increasingly common (e.g., Althuis et al., 2005; Atkin et al., 2008; Burgess et al., 2009). While breast cancer incidence, mortality, and survival rates differ by ethnicity, mortality rate reduction is associated with early detection and screening practices (Kudadjie-Gyamfi & Magai, 2008). This disease has become a serious public health concern especially in developing countries, as the application and effectiveness of early detection and diagnosis programs continue to lag (Modeste, Caleb-Drayton & Montgomery, 1999).

Latin Americans are widely impacted by breast cancer, as it is the leading cause of cancer death among this ethnicity (Abraido-Lanza, Chao & Gammon 2004). Breast cancer research in the United States, especially focused on Latin Americans, has become of primary importance, as these women tend to be diagnosed at more advanced stages than any other ethnic group (Hubbell, 2006). Thus, although Latin American women demonstrate a lower incidence of breast cancer, they are more likely than non-Hispanic White women to die from this disease (Lantz et al., 2006). Several studies demonstrate a higher death rate among Hispanic patients when compared to White women, despite the lower incidence of this minority in acquiring the disease (Holt, Lee & Wright, 2008). While health disparities and low health literacy act as barriers to Latin American women's breast cancer screening compliance, various other cultural and socioeconomic factors interact with this issue.

Venezuela, in particular, has a high level of incidence of women's breast cancer, a high mortality rate, and a high number of "years of potential productive life lost" (e.g., International Agency for Research on Cancer, 2008; MPPS, 2008). Although the government has a National Cancer Program that includes the breast cancer sub-program (Nigenda et al., 2011), the prevalent opinion is that the Venezuelan State has neglected this disease and its overall impact on Venezuelan women and their families (Sociedad Venezolana de Mastología, 2010). When compared with other Latin American countries, Venezuela presents some weaknesses (e.g., the absence of a national policy on breast cancer and the lack of a national breast cancer registry program), even though the country has made progress in the incorporation of civil society actors, including non-governmental organizations (NGOs), in the design and implementation of early detection programs (Nigenda et al., 2011).

This paper aims to discuss issues surrounding women's breast cancer such as awareness of the disease and its risks, self-advocacy, and the roles of activists, healthcare providers, and society. In particular, the paper aims to describe an action-oriented research project[1] that we developed in Venezuela about women's breast cancer, which was conducted over a four-year period of collaborative work among researchers, practitioners, NGOs, patients, journalists, and policymakers.[2] It discusses three key objectives of this project: 1) to facilitate the development of advocacy and spokesperson capabilities of survivors and community activists in order to increase their impact in the fight against women's breast cancer in Venezuela; 2) to contribute to closing the gap in the communication between physicians and patients and enhance the conversation between physicians, patients, and activists to define common goals to promote public awareness and advocacy; and 3) to promote the social dialogue among different stakeholders, including physicians, patients, activists, scholars, and public officials, in order to draft the foundations of a breast cancer national policy.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

AWARENESS OF WOMEN'S BREAST CANCER

Several studies indicate lagging breast health and cancer knowledge among various ethnic groups (Ott, Ullrich & Miller, 2009). A lack of awareness is a common problem that prevents women from engaging in productive breast health behaviors. Lagging awareness about this disease among Latin Americans is fuelled by culturally based misconceptions; however, while some women accurately understand its etiology, they misperceive their individual predispositions to breast cancer (Chavez et al., 1995). Therefore, understanding the *risk factors and predictors*, recognizing and practicing *early detection* measures, getting help from *supporting groups*, benefitting from *eHealth* and *awareness initiatives* are of great significance for women toward enhancing their awareness in the fight against women's breast cancer.

There is no concrete formula to developing or avoiding breast cancer (Canadian Cancer Society, 2004); however, the biggest risk factors are gender (i.e., being female), increased age, family history, and breast density (Breast Cancer Network of Strength, 2008). Others include: lifestyle habits, such as physical activity, diet, and weight (e.g., Courneya et al., 2006; Harvard Medical..., 2007); environmental determinants such as exposure to pollution, pesticides, and tobacco smoke (Plant, 2000); and hormonal factors, such as early onset of hormones, artificial hormones, and reproductive history (Cuzick, 2008). Due to the multitude of risk factors (Eid, 2003), failure to practice breast cancer prevention strategies is considered a risk in itself.

Various interventions and campaigns strive to raise awareness about breast cancer prevention, as early detection through breast self-examination or mammography is a woman's best chance for survival against the disease. If diagnosed, negative emotions (e.g., anxiety, depression, suicidal ideation, insomnia, and paranoia) can occur, inducing a variety of adjustment problems (Hawkins et al., 2008). Support groups for those facing breast cancer treatments provide an opportunity to talk about their feelings and experiences and obtain information about their diagnosis and treatment (Stang & Mittelmark, 2010).

Patient support resources have become increasingly integrated into cancer centers as more compassionate approaches to healthcare become acceptable and desirable (Eaton et al., 2005). Psychosocial oncology is a branch within breast cancer that deals with the psychological and social aspects of a breast cancer diagnosis. Counselors join forces with oncologists, nurses, and other healthcare professionals to create a cohesive breast cancer team, which focus specifically on the emotional and psychological impact of cancer (Kapusta & Wilcox, 2003). Psychosocial oncology also counters the hierarchical relationships that exist between medical experts and patients, which can elicit a heightened loss of control, illusion of incompetence, and a sense of meaninglessness for people who are in a chronic patient mode (Arntson & Droge, 1987). However, more efforts by medical professionals are required to acknowledge and refer patients to psychosocial branches of breast cancer treatment, as research demonstrates that people who use such resources spend less time with their clinicians, which results in decreased patient care costs and increased patient empowerment (Burns, 2005).

The emergence of eHealth—defined as the use of information and communication technology, especially the Internet, to improve or enable health and healthcare—enables greater access to healthcare information, provides an alternative to breast cancer support groups, and allows for a virtual exchange of ideas and information that is not confined by geography (Ginossar & Nelson, 2010). The Internet has allowed for a rapid increase in the formation of computer mediated support groups for people sharing any number of chronic or acute health conditions. To understand this phenomenon, Shaw and colleagues (2000) explore the Comprehensive Health Enhancement Support System (CHESS)—an interactive com-

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

puter program providing information, social support, and decision-making assistance to women with breast cancer. Their research indicates the usefulness of this program, demonstrated by the satisfaction, interactive productivity, and knowledge seeking behaviors exhibited by CHESS users.

While the Internet fosters growing breast cancer awareness, activism initiatives have transformed the ways in which this disease is understood, researched, and experienced (Moffett, 2003). Within the communities of people involved in raising awareness about this disease, social cohesion has emerged as an important force (e.g., Brody, Tickner & Rudel, 2005; Herstein, Mitki & Jaffe, 2008). Cohesive groups consist of members who demonstrate a desire to participate in a cause, respect for other group members, and are loyal and united in the pursuit of group's goals (Figueroa et al., 2002). Thus, as seen in most awareness initiatives, community and social cohesion is necessary for the success of campaigns and efforts with regards to the breast cancer cause (Harvey & Strahilevitz, 2009).

SELF-ADVOCACY

The notion of self-advocacy can benefit from the *patient-centered care* model for healthcare communication, *self-efficacy*, *health promotion*, and *coping strategies*. Various studies focus on patient-centered care[3] as the preferred model for healthcare communication (e.g., Albrecht et al., 2009; Bechtel & Ness, 2010; Clayton, Dudley & Musters, 2008; Hay et al., 2009; Kerr et al., 2003). This framework strives to decrease the power divide between doctors and patients, which is achieved through a dynamic process in which both parties understand one another as equals and experts in their own fields (Briggs & Mantini-Briggs, 2009). The idea of a mutual exchange is central to this method, which involves shared decision-making between the patient and physician and the establishment of a common ground and trust (Nekhlyudov & Braddock, 2009). Shared decision-making allows for a relationship of "collaborative informed choice" (DiMatteo, 1994, p. 155), which instigates increased patient satisfaction, as patients tend to feel intelligent, valued, and empowered (e.g., Ellingson & Buzzanell, 1999; Street et al., 2009). Although patient-centered care is often favored, precautions are still necessary. This paradigm requires a patient's willingness to participate in a specific manner, which can be seen as a challenge that may induce increased stress and anxiety in patients who already feel overwhelmed by their situation (e.g., Step et al., 2009; Swenson et al., 2004).

Patient support and empowerment[4] is a key priority of physician-patient communication. The role of information is absolutely crucial in patient empowerment (Brashers, Haas & Neidig, 1999); however, physicians are very much responsible for ensuring that patients receive the necessary information specific to their diagnosis (Luker et al., 1995). Patient-empowering communication on the part of the physician has been statistically and significantly linked to increased breast cancer knowledge among women, increased breast cancer screening rates, and higher quality of life for patients undergoing breast cancer screening and treatment (Maly et al., 2008). However, information dissemination must be culturally sensitive to effectively empower patients (Mishra et al., 1998).

The concept of self-efficacy refers to an individual's perceived control over their quality of health. A determinate of self-efficacy in changing behavior is the ability or perceived ability in maintaining behavioral changes; therefore, if individuals do not believe in their ability to maintain behavior changes, they are unlikely to implement these changes (Bandura, 1997). Other factors that affect self-efficacy are cultural precepts, psychosocial dispositions (e.g., depression and anxiety), and mistrust of medical systems (King & Wong, 2008). Self-efficacy is linked to education, knowledge, and empowerment,

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

which can result in healthy breast-screening behaviors (Rutten, Squiers & Hesse, 2006). Therefore, due to limited accessibility to healthcare institutions in countries where breast cancer awareness is limited and technology accessibility is limited or not up-to-date (Witt, 2008), it is increasingly important that communication programs capitalize on the usefulness of messages that instigate self-efficacy (Buki et al., 2004).

Health promotion is defined as "the process of enabling people to increase control over, and to improve, their health" (WHO, cited in Herbison & Lokanc-Diluzio, 2008, p. 303). In the context of women's breast cancer, it has become increasingly important to develop sound promotional strategies to encourage attention to breast health due to the high mortality rates associated with this disease. Communication strategies looking to persuade women must identify and address the deep structure of a woman's psyche, which involves the cultural, social, historical, and environmental factors that influence opinions and ideas about breast cancer (Barg & Grier, 2008). Medical empowerment is one of the best defenses available against the disease; therefore, effective messaging strategies are considered vital in the improvement of public health (Frisby, 2002).

Health promotion strategizing requires attention to the various ways in which women cope with breast cancer. While different people employ different coping mechanisms when faced with this disease, the most prominent methods are related to religion and spirituality, social support, and knowledge. Many females battling breast cancer seek religious and spiritual ventures, which are associated with worship and social activities that provide support, strength, and introspective guidance (Barg & Grier, 2008). Social support can also occur beyond the confines of organized religion, which is considered equally beneficial for those coping with the realities of this disease (Kapusta et al., 2005). The function of social support in helping women cope with breast cancer is the role of social relationships and the transaction of supportive behaviors among people, such as emotional, instrumental, informational, and appraisal support (Messina et al., 2004). Information "plays a key role in helping patients understand their disease process... enabling them to make informed medical decisions" (Talosig-Garcia & Davis, 2005, p. 53). Obtaining and processing information can increase autonomy and competence among breast cancer patients, as it provides women with a sense of control and empowerment in their treatment journey (Gustafson et al., 2008).

THE ROLE OF ACTIVISTS

Activism in Latin America is fundamental in the fight against women's breast cancer. It is achieved through the *NGO involvement*, the use of *research as an intervention tool*, and the processes of *networking* and *consensus building*. In virtually every Latin American country, some form of constitutional provision validates health as a basic human right. Despite this, the right to health is not guaranteed—it is a goal that is not legally enforceable. Thus, NGOs and community-based programs compliment such deficits, which are driven by social activism and guided by the values of collectivism found in Latin American culture. Health NGOs in Latin America demonstrate positive impacts on the long-term viability and welfare of the communities they operate in; they are considered a part of health activism efforts, which include a variety of roles (Zoller, 2005). These assemblies address some of the economic and social disparities that commonly prevent healthcare equality, which work to provide health, wellness, and lifestyle services to individuals and communities in need.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Due to the large social and economic disparities present within the healthcare system in much of Latin America, the efforts of NGOs have become increasingly crucial; however, both benefits and barriers require acknowledgement. One of the most beneficial aspects of NGOs becoming involved in health processes is their ability to create communicative spaces where the public can engage in education and deliberation about pertinent health issues (De Souza, 2009). Cancer-related NGOs have shown promise in being able to reach communities underserved by the public healthcare system, promoting the early detection of breast cancer (Durstine & Leitman, 2009).

However, some limitations to such involvement exist. With civil society involvement and community-based research that interlink with many actors and variables, there is ample opportunity for political social dynamic clashes (Israel et al., 1998). Additionally, many cancer NGOs are typically too small to enable measurable impacts in Latin America (Durstine & Leitman, 2009). Typically, cancer NGOs also do not collaborate with other NGOs, community organizations, and interested parties, which can diminish their ability to enact policy change. Due to these factors, NGOs are unlikely to fully reach target populations or provide them with the necessary prevention and early detection information.

The development and dissemination of knowledge is integral to effective interventions that strive to instigate healthy behaviors among populations. Health behavior theory is "based on the underlying notion of rational thought in which patients are assumed to be highly motivated to maintain or to return to a universally agreed upon definition of good health" (Barg & Grier, 2008, p. 336). That is, people make decisions about their health based on maximizing benefits and minimizing losses or costs. The theory of planned behavior is also relevant, which "hypothesizes that behavior is a product of three sets of independent beliefs—behavioral, control, and normative—which lead to the formation of attitudes; perceptions of behavior control; and subjective norms, respectively" (Courneya et al., 2006, p. 259).

The model of Communication for Social Change, which refers to an iterative process where community dialogue and collective action synergize to produce social change in a community, improving health and quality of life for all its members is also useful (Byrne et al., 2005). Similarly, the Integrated Model of Communication for Social Change refers to a "dynamic, iterative process that starts with a 'catalyst/stimulus' that can be external or internal to the community" (Byrne et al., 2005, p. 6). Perhaps the most important aspect of this model lies in the goal of knowledge equity; substantial emphasis is placed on the extent to which the knowledge is shared within a community rather than being hoarded or monopolized by just a few.

Research as an intervention tool is integral in healthcare. Focus group discussions (FGDs) in healthcare research are useful for obtaining ideas and perceptions on specific areas of interest in an open, permissive, and nonthreatening environment, which is especially beneficial when the topic of investigation is sensitive in nature (Ferrell et al., 1997). This research method is advantageous because it does not require a high degree of literacy or comprehension; thereby, it encourages participation from a wide demographic (e.g., OMNI Institute, 2004; Slaughter et al., 1999). Successful FGDs regarding breast cancer facilitate a forum that is open to varying backgrounds and beliefs, value all ideas and opinions, and foster intimate and attentive interaction (e.g., McCoy, 2003; National Breast and Cervical..., 2009, January 7). The benefit of using FGDs to address further concerns regarding breast cancer lies in their ability to elicit opinions, perceptions, attitudes, and concerns.

Effective networking exists in the realm of healthcare, thereby producing a wider reach of programs to various communities. Assessment, networking, and awareness of breast cancer incorporate a community perspective where those at the core of this activism are directly involved in program implementation. Participation in health and development communication programs not only strengthens the voice of

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

citizens but also ensures their involvement in decisions that affect them, their families, and their communities (Brasington, Tapia & van Lith, 2007). This approach favors mutual learning rather than top-down practices to change implementation, thereby respecting local traditions and refusing to undermine cultures and values of affected communities. However, careful consideration and planning is required to carry out network relations effectively; network integration must be intensive, involving multiple and overlapping links both within and across the organizations that compose the core of a network (Provan & Sebastian, 1998).

Identification of the facilitative conditions necessary to motivating breast cancer screening among women is paramount to increasing survival rates of this disease. Various studies emphasize the value of social network interventions as a powerful persuasion tool for influencing breast cancer screening among women (Volkman & Silk, 2008). Social connections can facilitate health behaviors due to the influences of naturally occurring social systems, such as family and friends (Allen, Stoddard & Sorensen, 2008). For example, women who engage in conversation about mammography with friends are more likely to get mammograms (Husaini et al., 2001). Therefore, understanding the relationship between social network characteristics and screening practices is vital in the improvement of lagging participation.

THE ROLE OF HEALTHCARE PROVIDERS

Healthcare providers play an essential role in the fight against women's breast cancer through the processes of *intervention* and *communication*. Interventions led by healthcare providers are considered an effective method in increasing breast cancer screening compliance among women (Rogers, Thompson & Robinson, 2002). The positive influence of healthcare providers in motivating women to engage in breast cancer screening is consistently demonstrated in literature examining this correlation. Perhaps the most important aspect of this linkage is the vital role of interpersonal communication (Dakroury & Eid, 2012) that can occur between a woman and a healthcare provider (Jones, Denham & Springston, 2006). While this area of communication offers hope for creating effective messaging strategies to encourage breast cancer screening and prevention, it includes confounding variables and complex issues. Overall, healthcare providers have the ability to educate women and motivate them to take control of their breast health. Therefore, it is important to understand the relational dynamics involved in this correlation to effectively harness the immense potential of these tendencies.

Healthcare practitioners must be aware of their vital function and be trained to interact proficiently with patients[5] to maximize the effectiveness of their messaging techniques (Diefenbach et al., 2009). Effective communication by physicians results in a plethora of positive effects, increasing patients' satisfaction of the healthcare experience and their overall health outcomes (Robinson et al., 2008). Communication that fosters the idea of humanism and an empathetic approach that treats patients as complete human beings with emotions and social needs is essential to effective physician-patient interactions (Butler et al., 2005).

Physicians play a pertinent role in the dissemination of information about breast cancer to patients (Lewis et al., 2009). Therefore, physicians need to focus on presenting clear and accurate information to aid the patient through their journey and instigate a relationship of trust, which is tied to an increase in compliance with physicians' recommendations (e.g., Kowalski et al., 2009; Paulsel, McCroskey & Richmond, 2006). Physicians are expected to recognize the contingent nature of medical visits, as patients have different informational needs at different times; physicians can address these needs in a way that avoids professional jargon[6] and is accurately understood by patients (e.g., Eheman et al., 2009;

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Thompson, 2009). Increased agency through decision-making, facilitated by the provision of information from physicians is beneficial for patients, as it is associated with increased participation, self-efficacy, and heightened trust (e.g., Arora et al., 2009; Vogel, Leonhart & Helmes, 2009).

Serious concerns are often associated with a diagnosis of breast cancer, necessitating strong physician-patient communication (Carney et al., 2009). Women tend to be more comfortable dealing with female physicians, especially when dealing with reproductive health issues, such as breast cancer, because of the perception of sharing a common female experience (Borrayo, Buki & Feigal, 2005). Women desire emotional support from their physicians more so than men, especially if they do not have strong familial support at home (Han et al., 2005). Physician-patient communication in this context often faces heightened complexities in developing countries, as resources are limited, breast health literacy is lagging, and access to properly trained staff and equipment can be challenging (Kreps & Sivaram, 2008).

Cultural competency in physician-patient communication is positively associated with improving women's abilities to make decisions regarding cancer treatment (Hawley et al., 2008). When dealing with Latin American women and breast cancer, physicians need to be actively aware of their patients' background, beliefs, and culture (Meeuwesen et al., 2007). Latin American women are less likely to trust their physicians as reliable and credible sources of information; therefore, they are less likely to follow their recommendations regarding screening practices (Otero-Sabogal et al., 2003). Especially in the Latin American context, the training of physicians may benefit from a family-centered rather than individual patient-centered approach to remain in alignment with the emphasis placed on family in Hispanic culture (Beckman & Frankel, 2003).

THE ROLE OF SOCIETY

Society has crucial impacts on patients in their experiences with breast cancer through two main influences—*familial* and *cultural*. Familialism is a crucial value in Latin American culture; it is the main source of support, especially in the presence of illness (Granda-Cameron, 1999). One of the most prominent roles that the family takes on is to provide emotional and social support[7] for female cancer patients, which can help to reduce the many negative psychological and psychosocial impacts of a breast cancer diagnosis (Nápoles-Springer et al., 2009). Loved ones are often the primary source of hope, encouragement, and emotional support; without this social support, the trauma women face may be devastating (Harris et al., 2009). Therefore, diagnosis of cancer does not only affect the patient; it affects the entire family.

This is especially true in Latin America where the protection of family members from harm is a strong belief (Mystakidou et al., 2004). Family also plays a role in how patients seek out and interpret information. Women often receive information about breast cancer from their loved ones (Washington et al., 2009). Women tend to feel as though they, not the physician, bear the responsibility to inform family members of their genetic cancer risk, which may also increase anxiety about a diagnosis (Macdonald et al., 2008). However, although the family may be the primary support-providers for women, this does not diminish the physician's responsibility to ensure that women are receiving support, whether from the family, the physician, social support groups, or online support groups (Levy & Strombeck, 2002).

Culture can seriously influence one's decision-making with regards to health behaviors. Factors such as religion and belief systems are woven into perceptions about healthcare, and thus are considered very important when looking to understand barriers related to proper breast cancer screening. An examination of Latin American cultural interactions with breast cancer reveals the prominence of fear and anxiety

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

related to the disease. That is, Latin American women fail to engage in screening and prevention practices due to the secretive and shameful perception of breast cancer, which is fuelled by embarrassment and shame likely associated with their limited knowledge of the disease (Hedeen & White, 2001). The lack of information exchange occurring at home is instigated by anxiety about discussing women's bodies—as this topic of conversation is considered private and inappropriate (Buki et al., 2004).

Religion and spirituality interact with Latin American women's views on breast cancer screening and treatment (Holt, Lee & Wright, 2008). While church-based health communication may not be at the forefront of screening compliance strategizing, it is considered an important tactic for some niche demographics that place a strong emphasis on the synergy between health and faith (Frisby, 2006). For instance, in their analysis of the effectiveness of a spiritually based approach to education on early detection of this disease, Holt, Lee, and Wright (2008) found that spiritually based interventions could be highly effective if properly strategized. Thus, interventions and communication techniques that demonstrate sensitivity to religious precepts can facilitate increased compliance to proper breast health practices.

ACTION INITIATIVE IN VENEZUELA

Action-oriented research is a collective, self-reflective, and participatory enquiry undertaken by participants (stakeholders) in social interactions to improve some condition in which they are all involved. It is "a methodological approach for doing collaborative research with practitioners and community partners that can inform practice, programs, community development, and policy while contributing to the scientific knowledge base" (Small & Uttal, 2005, p. 936). It promotes a critical consciousness that exhibits itself in political and practical action to make some type of change. Its agenda mainly includes "producing research that can address practical concerns" (Small, 1995, p. 942), aiming to "have some type of change" (Small & Uttal, 2005, p. 938). It helps to: 1) increase the closeness between the day-to-day problems encountered by practitioners in specific settings and the theories used to explain and resolve the problem; and 2) assist practitioners in lifting their veil of clouded understandings to better understand fundamental problems by raising their collective consciousness (Berg, 2009; Eid, 2011).

Action-oriented researchers "believe that the process of conducting research should be empowering for research participants" (Esterberg, 2002, p. 141). They "value collaboration with nonresearcher participants" in that while they bring "to the research process theoretical knowledge, experience, and the skills of conducting social science research, the participant collaborators bring practical knowledge and experience about the situations that are being studied" (Small, 1995, p. 942). Together, researchers and community partners "combine their different kinds of knowledge and skills to produce insightful and usable findings" (Small & Uttal, 2005, p. 938). Ultimately, action-oriented research aims to "generate knowledge that can be used to address practical concerns of local communities, organizations, and groups and incorporate local understandings of specific practices and issues" (Small & Uttal, 2005, p. 938).

In Venezuela, we developed an action-oriented research project in collaboration with SenosAyuda, a Venezuelan NGO that supports women with breast cancer and advocates for the early detection of the disease and better access to healthcare services, as well as a group of physicians, patients and survivors, journalists, and policymakers. This project aimed at helping SenosAyuda in the creation of a national network of breast cancer volunteer organizations while facilitating the acquisition of communication and advocacy skills and capabilities of breast cancer survivors and community activists. Our research has been guided by the "ecological perspective" (Street, 2003) that proposes an integration of the dif-

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

ferent levels in health communication (interpersonal, organizational, and social), and highlighting the potential emancipation role of health communication according to the pioneer of the field, Luis Ramiro Beltrán (1994).

The project was implemented in a very particular political and social environment. Since 1999, when Hugo Chavez became president after winning a democratic election, Venezuela has lived under a highly and virulent social and political polarization. The divide between the Chavez's followers and the opposition has translated into political violence, attacks to the freedom of speech, and a discourse of hate from both sides. However, the Chavez regime has put a strong emphasis on social programs in order to help the poor and expand the access to services, including access to healthcare, specially through its flag program *Misión Barrio Adentro* (Inside the Neighborhood Mission) with the help of the government of Cuba (Briggs & Mantini-Briggs, 2009). The overall performance of the *Misión* is far from the positive results announced by the government; this program has even contributed to an increase in the historical fragmentation of the Venezuelan healthcare system with a negative impact on the overall financing of the public sector (e.g., Díaz Polanco, 2006; Nahon-Serfaty, 2005). Social challenges in Venezuela include general resistance from the public to learn about women's breast cancer, shortage in physicians' availability, lack of training of spokespeople on behalf of patients, survivors, and foundations, shortages in the logistics of organizing public events, absence or reluctance of volunteers and advocates, and deficiencies in activity evaluation and feedback measures.

The civil society organizations emerged in Venezuela as a way to fill the void left by the public sector in the provision of healthcare, especially to the poor populations. The oldest organization, the *Sociedad Anticancerosa de Venezuela* (the Cancer Society of Venezuela) was created in 1948 to lead the efforts to educate and provide medical care to those lacking financial means or insurance coverage. Recently many NGOs working specifically in breast cancer have been created: SenoSalud, Funcamama, Fuccam, Famac, Fundamama, Fundaseno, Senosestima, Unicam, and SenosAyuda, among others.

SenosAyuda was created to address the problems and deficiencies that women are facing with breast cancer issues, and promote nationwide long-term action programs. In 2007, the NGO initiated the Organizational Capacity Building Program (OCBP) directed to collaborate with other social development organizations involved in the fight against women's breast cancer in Venezuela. The OCBP objectives were: 1) to contribute to the development of such organizations in three fundamental phases of their evolution: to be born and established, to progress, and to be consolidated or redefined; and 2) to educate the leadership hands-on learning methodologies and provide the social development organizations with the managerial tools that allow them to manage their associations efficiently.

Within the context of the OCBP initiative, our project was based on six guiding principles: *inclusion and equity*, *gender sensitivity*, *dialogue and interaction*, *social change*, *networking*, and *ethical standards*. Inclusion and equity are fundamental because all the stakeholders should have a voice in defining their priorities and needs, and all have the potential of becoming advocates and spokespersons of their own cause. Gender sensitivity takes into account that on top of all the traditional barriers to access proper healthcare services, women in Venezuela face additional obstacles due to cultural and social obstacles as they are frequently the subject of a double marginalization process (being poor and women). Dialogue and interaction are ways to promote an open conversation about the goals, strategies, and tactics of the project, inspired by the conviction that communication is a two-way and horizontal process (e.g., Eid & Dakroury, 2010). Social change is understood here as a gradual process that entails the development of capabilities and abilities of the individuals and organizations to improve their access to information, healthcare services and participation in the policy decision-making. Networking is a way to promote

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

solidarity among different actors, including those who were marginalized (i.e., regional activists) and to create synergies among them. Ethical standards are essential to respect human dignity, promote transparency, encourage responsibility, and demand accountability of all the stakeholders (e.g., Eid, 2008; Eid & Nahon-Serfaty, 2008).

Collaborators of this action-oriented research project have engaged in discussions through four annual collective workshops in Caracas and continuous national events and activities throughout the four-year period (2009-2012). The workshops essentially helped participants to be spokespeople for their organizations and to understand how to use communication as a tool to plan and achieve the given activities. The workshops provided skills of composing and communicating messages about breast health that help promote changes in attitude and behavior. They discussed the situation of patients living in regions with no access to healthcare services, the confusion caused by the debate about breast implants, and the need to secure humane and dignified palliative care. They helped also in how to determine the importance of having volunteers and thanking them for their contributions, use techniques of group discussions, target the audiences and select the strategies to reach them, and conduct social mobilization. Social mobilization involves how to work together to create visibility of the problem and achieve support of the government, opposition members of the council, and community members.

Events and activities were numerous. Collaborators have been active in conducting and participating in events and activities in order to promote breast cancer health awareness. October 19th was the day dedicated to promoting breast health. In particular, "The Pink Walk" was an important event where breast cancer survivors and other activists from the community could walk the streets, raising awareness. In addition, posters and banners were put up in public places such as bus stops and taxi and moto-taxi stops. Activities include the production of materials, such as the "Pink Recipe," which discusses the four "to-dos" for breast health: self-exam, breast echography, mammogram, and consultation with a physician. Other materials were newsletters, letters asking for contributions and thanking sponsors and volunteers, and website updates and maintenance. Essentially, the most significant activities involved getting the messages of breast health out to the public, the media, and even physicians. The publics were reached through direct talks (such as the ABC's of breast cancer—talks about the organization, activities, and early detection and prevention), distribution of pamphlets, information, and pink ribbons. For example, the talks with the Mission Mothers of Barrio helped reach women in poorer communities. The creation of clubs called, "Prevention and Early Diagnostic of Breast Cancer", involved different talks about breast health, the creation of billboards and newsletters, and even theatre presentations.

One of the major outcomes of our action-oriented research project has been the expansion of the networking base of SenosAyuda with different stakeholders including other NGOs from different regions, physicians, scholars, journalists, and public officials from state and municipal governments.[8] For example, a direct outcome celebrated in May 2011 after the third workshop of our project was the declaration "*For a Consensus Vision for the formulation and implementation of a National Response against Breast Cancer in Venezuela*" which was subscribed by an initial group of 20 participants. A year after, in May 2012, the document has been signed by 101 organizations, including 8 breast cancer advocacy organizations, 6 major medical associations, 4 human rights NGOs, and important social actors, such as the Red Cross and the National College of Journalists.

The declaration has also contributed to set additional goals of networking and consensus building for SenosAyuda, the organization that is leading the initiative, including the collection of 10,000 signatures by the end of 2013 in order to support a National Response against Breast Cancer in Venezuela, the launch of national media campaign to raise public awareness about the need of a consensus, and the

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

celebration of public forum all over the country to discuss the principles and objectives that should guide a national policy for the disease. The declaration has also served as a framework to define the priorities of a future national policy on breast cancer:

- Educate and train healthcare personnel specialized in the diagnostic, treatment, and integral support of breast cancer patients;
- Develop a national system to compile, analyze, and facilitate the access to reliable data that show the real impact of breast cancer on the population as a way to draft evidence-based policies and actions;
- Establish the mammography as a key procedure for an accurate diagnosis that also contributes to articulate the different players in the healthcare system;
- Communicate messages associated with the solutions available to the public and adapted to their needs and socio-cultural realities;
- Secure financial protection to every person affected by breast cancer through public and private healthcare insurances and funding;
- Promote the integration of public and private services in order to improve access to proper diagnosis procedures and treatments at the national level;
- Provide palliative care as part of human, moral, and ethical duties towards people suffering from breast cancer.

CONCLUSION

Breast cancer continues to plague female populations around the globe, necessitating awareness initiatives designed to increase participation in prevention techniques, such as screening and mammography. The prevalence of women's breast cancer in Venezuela is particularly alarming, as it is attributed to healthcare inequalities, low health literacy, and lagging compliance with prevention methods. While the right to health is acknowledged by the Venezuelan constitution, activism beyond governmental confines is required to increase breast cancer awareness and decrease mortality rates. Although a simple recipe has yet to be discovered that results in breast cancer, a wide variety of risk factors and predictors are involved in breast cancer diagnoses. However, through social interactions facilitated via support groups and online forums, it is possible for women to exchange ideas and knowledge regarding this disease. In addition, awareness initiatives organized and facilitated by a variety of groups aid in women's breast cancer consciousness-raising efforts.

Individuals who face the burden of breast cancer are bombarded by a variety of factors that interact with one's disease path. Patients are forced to wade through their own emotions and those of others, information regarding treatment, and a multitude of healthcare practitioner personalities. While this process can be extremely daunting for many women, the enactment of certain behaviors and social mechanisms can act as useful coping techniques. These tactics are intrinsically linked to empowerment and are centralized around one's ability to enact self-advocacy, which are facilitated through patient-centered care, self-efficacy, and strategic health promotion and coping strategies. Such empowerment can boost feelings of self-efficacy, which may promote healthy preventative behaviors (Errico & Rowden, 2006). Particularly in Latin America, NGOs are involved in the instigation of these endeavors; however, they commonly face obstructions to their success. Thus, research surrounding health behaviors is identified as

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

a useful tool in guiding health interventions. The techniques and ideologies extracted from such research are subsequently complimented by networking, which fosters interactions and community involvement necessary for breast cancer awareness among women.

Activism initiatives function with the help of both healthcare providers and social influences. Healthcare providers play a unique role in activism and raising awareness, as they have the ability to interact with patients on a personal and intimate level. However, this opportunity for enacting satisfaction and healthy behaviors among patients relies on healthcare providers' communicative abilities. A detailed analysis of physician-patient communication reveals the highly complex nature of such interactions and the need for ongoing training and development. The role of society, specifically familial and cultural influences, has a strong impact on women's breast cancer awareness, ultimately affecting how breast cancer patients choose to navigate their healing journey. In fact, Latin American women respond more positively to physicians and outreach that is sensitive and attuned to their specific cultural context (e.g., Larkey, 2006; Warren, Londoño & Wessell, 2006).

As a result of the workshops, events, and activities of our action-oriented research project in Venezuela, there has been an increasing interest among community members to learn and seek more information, a higher level of awareness of the concerned organizations, a better understanding of the communicated messages, an increase in the media coverage and the number of medical consultations, and positive satisfaction reports about how patients were treated in the provided healthcare services. Breast cancer risk factors have been communicated through positive messages such as those of support, the colour pink, lifestyle changes, keeping health as a determining factor for life, fighting and winning life, and knowing your breasts. The decree of the Municipal Council includes data about mortalities from breast cancer, about access to healthcare services and risk factors, and of promoting October 19th as "Breast Cancer Day". The idea of solidarity and social responsibility in that health and well-being have an influence on families and communities was successfully communicated through the increase of awareness and promotion of social change.

REFERENCES

Abraído-Lanza, A. F., Chao, M. T., & Gammon, M. D. (2004). Breast and cervical cancer screening among Latinas and Non-Latina whites. *American Journal of Public Health*, *94*(8), 1393–1398. doi:10.2105/AJPH.94.8.1393 PMID:15284049

Albrecht, T., Penner, L. A., Cline, R. J. W., Eggly, S. S., & Ruckdeschel, J. C. (2009). Studying the process of clinical communication: Issues of context, concepts, and research directions. *Journal of Health Communication*, *14*(S1), 47–56. doi:10.1080/10810730902806794 PMID:19449268

Allen, J. D., Stoddard, A. M., & Sorensen, G. (2008). Do social network characteristics predict mammography screening practices? *Health Education & Behavior*, *35*(6), 763–776. doi:10.1177/1090198107303251 PMID:17620665

Althuis, M. D., Dozier, J. M., Anderson, W. F., Devesa, S. S., & Brinton, L. A. (2005). Global trends in breast cancer incidence and mortality 1973-1997. *International Journal of Epidemiology*, *34*(2), 405–412. doi:10.1093/ije/dyh414 PMID:15737977

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Arntson, P., & Droge, D. (1987). Social support in self-help groups: The role of communication in enabling perceptions of control. In T. L. Albrecht & M. B. Adelman (Eds.), *Communicating social support* (pp. 148–171). California: Sage.

Arora, N. K., Weaver, K. E., Clayman, M. L., Oakley-Girvan, I., & Potosky, A. L. (2009). Physicians' decision-making and psychosocial outcomes among cancer survivors. *Patient Education and Counseling*, *77*(3), 404–412. doi:10.1016/j.pec.2009.10.004 PMID:19892508

Atkin, C. K., Smith, S. W., McFeters, C., & Ferguson, V. (2008). A comprehensive analysis of breast cancer news coverage in leading media outlets focusing on environmental risks and prevention. *Journal of Health Communication*, *13*(1), 3–19. doi:10.1080/10810730701806912 PMID:18307133

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W.H. Freeman & Company.

Barg, F. K., & Grier, S. A. (2008). Enhancing breast cancer communications: A cultural models approach. *International Journal of Research in Marketing*, *25*(4), 335–342. doi:10.1016/j.ijresmar.2008.07.003

Bechtel, C., & Ness, D. L. (2010). If you build it, will they come? Designing truly patient-centered healthcare. *Health Affairs*, *29*(5), 914–920. doi:10.1377/hlthaff.2010.0305 PMID:20439880

Beckman, H. B., & Frankel, R. M. (2003). Training practitioners to communicate effectively in cancer care: It is the relationship that counts. *Patient Education and Counseling*, *50*(1), 85–89. doi:10.1016/S0738-3991(03)00086-7 PMID:12767591

Beltrán, L. R. (1994). La salud y la comunicación en Latinoamérica: Políticas, estrategias y planes. In *UNESCO - OPS, Por una política de comunicación para la promoción de la salud en América Latina* (pp. 29–90). Quito, Ecuador: OPS.

Berg, B. L. (2009). *Qualitative research methods for the social sciences*. Boston, MA: Allyn & Bacon.

Borrayo, E. (2005). Breast cancer detection among older Latinas: Is it worth the risk? *Qualitative Health Research*, *15*(9), 1244–1263. doi:10.1177/1049732305281337 PMID:16204403

Brashers, D. E., Haas, S. M., & Neidig, J. L. (1999). The patient self-advocacy scale: Measuring patient involvement in healthcare decision-making interactions. *Health Communication*, *11*(2), 97–121. doi:10.1207/s15327027hc1102_1 PMID:16370972

Brasington, A., Tapia, M. & van Lith, L. (2007). *Involving those directly affected in health and development communication programs.* Baltimore: Health Communication Partnership based at the John Hopkins Bloomberg School of Public Health/Center for Communication Programs.

Breast Cancer Network of Strength. (2008). *A woman's guide to breast care* [Brochure]. Chicago, IL, USA.

Briggs, C. L., & Mantini-Briggs, C. (2009). Confronting health disparities: Latin American social medicine in Venezuela. *American Journal of Public Health*, *99*(3), 549–555. doi:10.2105/AJPH.2007.129130 PMID:19150916

Brody, J. G., Tickner, J., & Rudel, R. A. (2005). Community-initiated breast cancer and environment studies and the precautionary principle. *Environmental Health Perspectives*, *113*(8), 920–925. doi:10.1289/ehp.7784 PMID:16079059

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Buki, L. P., Borrayo, E. A., Feigal, B. M., & Carrillo, I. Y. (2004). Are all Latinas the same? Perceived breast cancer screening barriers and facilitative conditions. *Psychology of Women Quarterly*, *28*(4), 400–411. doi:10.1111/j.1471-6402.2004.00157.x

Burgess, C. C., Linsell, L., Kapari, M., Omar, L., Michell, M., Whelehan, P., & Ramirez, A. J. et al. (2009). Promoting early presentation of breast cancer by older women: A preliminary evaluation of a one-to-one health professional-delivered intervention. *Journal of Psychosomatic Research*, *67*(5), 377–387. doi:10.1016/j.jpsychores.2009.01.005 PMID:19837200

Burns, K. (2005). *Focus on solutions: A health professional's guide*. London: Whurr Publishers.

Butler, L., Degner, L., Baile, W., & Landry, M. (2005). Developing communication competency in the context of cancer: A critical interpretive analysis of provider training programs. *Psycho-Oncology*, *14*(10), 861–872. doi:10.1002/pon.948 PMID:16200525

Byrne, A., Gray-Felder, D., Hunt, J., & Parks, W. (2005). *Who measures change? An introduction to participatory monitoring and evaluation of communication for social change*. UK: Department for International Development, the Rockefeller Foundation, and Communication for Social Change Consortium.

Canadian Cancer Society. (2004). Eat well, be active: What you can do. [Brochure]. Toronto, ON, Canada.

Carney, P. A., Kettler, M., Cook, A. J., Geller, B. M., Karliner, L., Miglioretti, D. L., & Elmore, J. G. et al. (2009). An assessment of the likelihood, frequency, and content of verbal communication between radiologists and women receiving screening and diagnostic mammography. *Academic Radiology*, *16*(9), 1056–1063. doi:10.1016/j.acra.2009.02.023 PMID:19442539

Cazap, E., Buzaid, A., Garbino, C., de la Garza, J., Orlandi, F., Schwartsmann, G., & Breitbart, G. et al. (2010). Breast cancer in Latin America: Experts perceptions compared with medical care standards. *The Breast*, *19*(1), 50–54. doi:10.1016/j.breast.2009.10.011 PMID:19945878

Chavez, L. R., Hubbell, A., McMullin, J. M., Martinez, R. G., & Mishra, S. I. (1995). Structure and meaning in models of breast and cervical cancer risk factors: A comparison of perceptions among Latinas, Anglo women, and physicians. *Medical Anthropology Quarterly*, *9*(1), 40–74. doi:10.1525/maq.1995.9.1.02a00030 PMID:7697550

Clayton, M. F., Dudley, W. N., & Musters, A. (2008). Communication with breast cancer survivors. *Health Communication*, *23*(3), 207–221. doi:10.1080/10410230701808376 PMID:18569050

Courneya, K. S., Jones, L. W., Mackey, J. R., & Fairey, A. S. (2006). Exercise beliefs of breast cancer survivors before and after participation in a randomized controlled trial. *International Journal of Behavioral Medicine*, *13*(3), 259–264. doi:10.1207/s15327558ijbm1303_10 PMID:17078777

Cuzick, J. (2008). Short communication: Assessing risk for breast cancer. *Breast Cancer Research*, *10*(4Suppl 4), S13. doi:10.1186/bcr2173 PMID:19128426

Dakroury, A., & Eid, M. (2012). Introduction to communication and media studies. In M. Eid & A. Dakroury (Eds.), *Basics in communication and media studies* (pp. 3–31). Boston, MA: Pearson.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

De Souza, R. (2009). Creating "communicative spaces": A case of NGO community organizing for HIV/AIDS prevention. *Health Communication*, *24*(8), 692–702. doi:10.1080/10410230903264006 PMID:20183378

Díaz Polanco, J. (2006). Salud y Hegemonía Regional: Las relaciones Venezuela-Cuba, 1999-2006. *Foreign Affairs en Español*, *6*(4), 10–12.

Diefenbach, M., Turner, G., Carpenter, K. M., Sheldon, L. K., Mustian, K. M., Gerend, M. A., & Miller, S. M. et al. (2009). Cancer and physician-patient communication. *Journal of Health Communication*, *14*(S1), 57–65. doi:10.1080/10810730902814079 PMID:19449269

DiMatteo, M. R. (1994). The physician-patient relationship: Effects on the quality of healthcare. *Clinical Obstetrics and Gynecology*, *37*(1), 149–161. doi:10.1097/00003081-199403000-00019 PMID:8194205

Durstine, A., & Leitman, E. (2009). Building a Latin American cancer patient advocacy movement: Latin American cancer NGO regional overview. *Salud Pública de México*, *51*(2), 316–322. doi:10.1590/S0036-36342009000800023 PMID:19967288

Eaton, G., Hryniuk, W., Kapusta, B., Savage, C., Wilcox, B., & Yurichuk, S. (2005). Report card 2004. *Cancer Advocacy Coalition of Canada*, *7*, 1–40.

Eheman, C. R., Berkowitz, Z., Lee, J., Mohile, S., Purnell, J., Rodriguez, E. M., & Morrow, G. et al. (2009). Information-seeking styles among cancer patients before and after treatment by demographics and use of information sources. *Journal of Health Communication*, *14*(5), 487–502. doi:10.1080/10810730903032945 PMID:19657928

Eid, M. (2003). Reflexive modernity and risk society. *The International Journal of the Humanities*, *1*(1), 813–828.

Eid, M. (2008). *Interweavement: International media ethics and rational decision-making*. Boston, MA: Pearson.

Eid, M. (2011). Introduction to communication research. In M. Eid (Ed.), *Research methods in communication* (pp. 3–14). Boston, MA: Pearson.

Eid, M., & Dakroury, A. (Eds.). (2010). *Communication and media studies: An introduction*. Boston, MA: Pearson.

Eid, M., & Nahon-Serfaty, I. (2008). Transparency in direct-to-consumer pharmaceutical marketing: Crucial ethical standards. *IMC Review: Journal of Integrated Marketing Communications*, *7*(1), 6–14.

Ellingson, L. L., & Buzzanell, P. M. (1999). Listening to women's narratives of breast cancer treatment: A feminist approach to patient satisfaction with physician-patient communication. *Health Communication*, *11*(2), 153–183. doi:10.1207/s15327027hc1102_3 PMID:16370974

Errico, K. M., & Rowden, D. (2006). Experiences of Breast Cancer survivor-advocates and advocates in countries with limited resources: A shared journey in breast cancer advocacy. *The Breast Journal*, *12*(1), S111–S116. doi:10.1111/j.1075-122X.2006.00208.x PMID:16430394

Esterberg, K. G. (2002). *Qualitative methods in social research*. Boston, MA: McGraw-Hill.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Ferrell, B. R., Grant, M. M., Funk, B., Otis-Green, S., & Garcia, N. (1997). Quality of life in breast cancer survivors as identified by focus groups. *Psycho-Oncology*, *6*(1), 13–23. doi:10.1002/(SICI)1099-1611(199703)6:1<13::AID-PON231>3.0.CO;2-S PMID:9126712

Figueroa, M.E., Kincaid, L.D., Rani, M., Lewis, G. & Gray-Felder, D. (2002). *Communication for social change: An integrated model for measuring the process and its outcomes*. Baltimore: Health Communication Partnership based at the John Hopkins Bloomberg School of Public Health/Center for Communication Programs and Rockefeller Foundation as part of their Communication for Social Change Grantmaking Strategy.

Frisby, C. M. (2002). Messages of hope: Health communication strategies that address barriers preventing black women from screening for breast cancer. *Journal of Black Studies*, *32*(5), 489–505. doi:10.1177/002193470203200501

Frisby, C. M. (2006). A matter of life and death: Effects of emotional message strategies on African American women's attitudes about preventative breast cancer screenings. *Journal of Black Studies*, *37*(1), 103–126. doi:10.1177/0021934705277499

Ginossar, T., & Nelson, S. (2010). La communidad habla: Using Internet community-based information interventions to increase empowerment and access to healthcare of low income Latino/a immigrants. *Communication Education*, *59*(3), 328–343. doi:10.1080/03634521003628297

Gonzalez, G., Gallardo, N., & Bastani, R. (2005). A pilot study to define social support among Spanish-speaking women diagnosed with a breast abnormality suspicious for cancer: A brief research report. *Journal of Psychosocial Oncology*, *23*(1), 109–120. doi:10.1300/J077v23n01_07 PMID:16492647

Granda-Cameron, C. R. N. (1999). The experience of having cancer in Latin America. *Cancer Nursing*, *22*(1), 51–57. doi:10.1097/00002820-199902000-00010 PMID:9990758

Gustafson, D. H., Hawkins, R., McTavish, F., Pingree, S., Chen, W. C., Volrathongchai, K., & Serlin, R. C. et al. (2008). Internet-based interactive support for cancer patients: Are integrated systems better? *Journal of Communication*, *58*(2), 238–257. doi:10.1111/j.1460-2466.2008.00383.x PMID:21804645

Han, W. T., Collie, K., Koopman, C., Azarow, J., Classen, C., Morrow, G. R., & Spiegel, D. et al. (2005). Breast cancer and problems with medical interactions: Relationships with traumatic stress, emotional self-efficacy, and social support. *Psycho-Oncology*, *14*(4), 318–330. doi:10.1002/pon.852 PMID:15386762

Hanoch, Y. (2004). Improving doctor-patient understanding of probability in communicating cancer-screening test findings. *Journal of Health Communication*, *9*(4), 327–335. doi:10.1080/10810730490468496 PMID:15371085

Harris, J., Bowen, D. J., Badr, H., Hannon, P., Hay, J., & Sterba, K. R. (2009). Family communication during the cancer experience. *Journal of Health Communication*, *14*(S1), 76–84. doi:10.1080/10810730902806844 PMID:19449271

Harvard Medical School's Center of Excellence in Women's Health. (2007). *Black women's attitudes about the links between breast cancer and obesity*. Boston, MA.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Harvey, J. A., & Strahilevitz, M. A. (2009). The power of pink: Cause-related marketing and the impact on breast cancer. *Journal of the American College of Radiology*, *6*(1), 26–32. doi:10.1016/j.jacr.2008.07.010 PMID:19111268

Hawkins, R., Hwang, H., Lee, S. Y., & Pingree, S. (2008). Interplay of negative emotion and health self-efficacy on the use of health information and its outcomes. *Communication Research*, *35*(3), 358–381. doi:10.1177/0093650208315962

Hawley, S. T., Janz, N. K., Hamilton, A., Griggs, J. J., Alderman, A. K., Mujahid, M., & Katz, S. J. (2008). Latina patient perspectives about informed treatment decision making for breast cancer. *Patient Education and Counseling*, *73*(2), 363–370. doi:10.1016/j.pec.2008.07.036 PMID:18786799

Hay, J., Harris, J. N., Waters, E. A., Clayton, M. F., Ellington, L., Abernethy, A. D., & Prayor-Patterson, H. (2009). Personal communication in primary and secondary cancer prevention: Evolving discussions, emerging challenges. *Journal of Health Communication*, *14*(1), 18–29. doi:10.1080/10810730902806828 PMID:19449265

Hedeen, A. N., & White, E. (2001). Breast cancer size and stage in Hispanic American women, by birthplace: 1992-1995. *American Journal of Public Health*, *91*(1), 122–125. doi:10.2105/AJPH.91.1.122 PMID:11189803

Herbison, S., & Lokanc-Diluzio, W. (2008). The novella approach to inform women living on low income about early breast cancer detection. *Health Promotion Practice*, *9*(3), 294–304. doi:10.1177/1524839906289166 PMID:16923845

Herstein, R., Mitki, Y., & Jaffe, E. D. (2008). Israel Cancer Association: Promoting the fight against breast cancer with IMC strategy. *International Journal of Pharmaceutical and Healthcare Marketing*, *2*(1), 63–75. doi:10.1108/17506120810865433

Holt, C. L., Lee, C., & Wright, K. (2008). A spiritually based approach to breast cancer awareness: Cognitive response analysis of communication effectiveness. *Health Communication*, *23*(1), 13–22. doi:10.1080/10410230701626919 PMID:18443989

Hubbell, A. P. (2006). Mexican American women in a rural area and barriers to their ability to enact protective behaviors against breast cancer. *Health Communication*, *20*(1), 35–44. doi:10.1207/s15327027hc2001_4 PMID:16813487

Huijer, M., & van Leeuwen, E. (2000). Personal values and cancer treatment refusal. *Journal of Medical Ethics*, *26*(4), 358–362. doi:10.1136/jme.26.5.358 PMID:11055039

Husaini, B. A., Sherkat, D. E., Bragg, R., Levine, R., Emerson, J. S., Mentes, C. M., & Cain, V. A. (2001). Predictors of breast cancer screening in a panel study of African American women. *Women & Health*, *34*(3), 35–51. doi:10.1300/J013v34n03_03 PMID:11708686

International Agency for Research on Cancer. (2008). *Globocan 2008: Fast stats*. Retrieved from http://globocan.iarc.fr/factsheet.asp#WOMEN

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Israel, B. A., Schulz, A. J., Parker, E. A., & Becker, A. B. (1998). Review of community-based research: Assessing partnership approaches to improve public health. *Annual Review of Public Health*, *19*(1), 173–202. doi:10.1146/annurev.publhealth.19.1.173 PMID:9611617

Jones, K. O., Denham, B. E., & Springston, J. K. (2006). Effects of mass and interpersonal communication on breast cancer screening: Advancing agenda-setting theory in health contexts. *Journal of Applied Communication Research*, *34*(1), 94–113.

Kapusta, B., Hryniuk, W., Savage, C., Eaton, G., Yurichuk, S., & Wilcox, B. (2005). Report card 2004. *Cancer Advocacy Coalition of Canada*, *7*(Winter), 1–40.

Kapusta, B., & Wilcox, B. (2003). *The emotional facts of life with cancer: A guide to counseling and support for patients, families and friends*. Toronto: Canadian Association of Psychosocial Oncology.

Kerr, J., Engel, J., Schlesinger-Raab, A., Sauer, H., & Hölzel, D. (2003). Doctor-patient communication: Results of a four-year prospective study in rectal cancer patients. *Diseases of the Colon and Rectum*, *46*(8), 1038–1046. doi:10.1007/s10350-004-7278-6 PMID:12907897

Kowalski, C., Nitzsche, A., Scheibler, F., Steffen, P., Albert, U., & Pfaff, H. (2009). Breast cancer patients' trust in physicians: The impact of patients' perception of physicians' communication behaviors and hospital organizational climate. *Patient Education and Counseling*, *77*(3), 344–348. doi:10.1016/j.pec.2009.09.003 PMID:19818577

Kreps, G. & Sivaram, R. (2008). Strategic health communication across the continuum of breast cancer care in limited-resource countries. *American Cancer Society, 113*(8), 2331-2337.

Kudadjie-Gyamfi, E., & Magai, C. (2008). The influence of coping styles on mammography screening in a multiethnic sample. *Cultural Diversity & Ethnic Minority Psychology*, *14*(3), 183–192. doi:10.1037/1099-9809.14.3.183 PMID:18624582

Lantz, P. M., Mujahid, M., Schwartz, K., Janz, N. K., Fagerlin, A., Salem, B., & Katz, S. J. et al. (2006). The influence of race, ethnicity, and individual socioeconomic factors on breast cancer stage at diagnosis. *American Journal of Public Health*, *96*(12), 2173–2178. doi:10.2105/AJPH.2005.072132 PMID:17077391

Larkey, L. (2006). Las mujeres saludables: Reaching Latinas for breast, cervical and colorectal cancer prevention and screening. *Journal of Community Health*, *31*(1), 69–77. doi:10.1007/s10900-005-8190-2 PMID:16482767

Lee, S. Y., & Hawkins, R. (2010). Why do patients seek an alternative channel? The effects of unmet needs on patients' health-related internet use. *Journal of Health Communication*, *15*(2), 152–166. doi:10.1080/10810730903528033 PMID:20390984

Levy, J. A., & Strombeck, R. (2002). Health benefits and risks of the internet. *Journal of Medical Systems*, *26*(6), 495–510. doi:10.1023/A:1020288508362 PMID:12385532

Lewis, N., Gray, S. W., Freres, D. R., & Hornik, R. C. (2009). Examining cross-source engagement with cancer-related information and its impact on doctor-patient relations. *Health Communication*, *24*(8), 723–734. doi:10.1080/10410230903264030 PMID:20183381

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Luker, K. A., Beaver, K., Leinster, S. J., Owens, R. G., Degner, L. F., & Sloan, J. A. (1995). The information needs of women newly diagnosed with breast cancer. *Journal of Advanced Nursing*, *22*(1), 134–141. doi:10.1046/j.1365-2648.1995.22010134.x PMID:7560521

Macdonald, D. J., Sarna, L., Giger, J. N., Van Servellen, G., Bastani, R., & Weitzel, J. N. (2008). Comparison of Latina and non-Latina white women's beliefs about communicating genetic cancer risk to relatives. *Journal of Health Communication*, *13*(5), 465–479. doi:10.1080/10810730802198920 PMID:18661388

Maly, R. C., Stein, J. A., Umezawa, Y., Leake, B., & Anglin, M. D. (2008). Racial and ethnic differences in breast cancer outcomes among older patients: Effects of physician communication and patient empowerment. *Health Psychology*, *27*(6), 728–736. doi:10.1037/0278-6133.27.6.728 PMID:19025268

McCoy, S. (2003). *Discussion guide for phase I focus groups (APHA): Perceptions of TB in the Hispanic community*. Atlanta: Centers for Disease Control and Prevention.

Meeuwesen, L., Tromp, F., Schouten, B. C., & Harmsen, J. A. M. (2007). Cultural differences in managing information during medical interaction: How does a physician get a clue? *Patient Education and Counseling*, *67*(1), 183–190. doi:10.1016/j.pec.2007.03.013 PMID:17467947

Messina, C. R., Lane, D. S., West, D. S., Frishman, W., Glanz, K., Taylor, V., & Powell, L. (2004). Relationship of social support and social burden to repeated breast cancer screening in the women's health initiative. *Health Psychology*, *23*(6), 582–594. doi:10.1037/0278-6133.23.6.582 PMID:15546226

Mishra, S. I., Chavez, L. R., Magaňa, R. J., Nava, P., Valdez, R. B., & Hubbell, F. A. (1998). Improving breast cancer control among Latinas: Evaluation of a theory-based educational program. *Health Education & Behavior*, *25*(5), 653–670. doi:10.1177/109019819802500511 PMID:9768384

Modeste, N. N., Caleb-Drayton, V. L., & Montgomery, S. (1999). Barriers to early detection of breast cancer among women in a Caribbean population. *Pan American Journal of Public Health*, *5*(3), 152–156. doi:10.1590/S1020-49891999000300003 PMID:10355312

Moffett, J. (2003). Moving beyond the ribbon: An examination of breast cancer advocacy and activism in the US and Canada. *Cultural Dynamics*, *15*(3), 287–306. doi:10.1177/09213740030153004

MPPS. (2008). *Anuario de Mortalidad*. Caracas, Venezuela: Ministerio del Poder Popular de la Salud.

Mystakidou, K., Parpa, E., Tsilika, E., Katsouda, E., & Vlahos, L. (2004). Cancer information disclosure in different cultural contexts. *Supportive Care in Cancer*, *12*(3), 147–154. doi:10.1007/s00520-003-0552-7 PMID:15074312

Nahon-Serfaty, I. (2005). *Le discours sur la réforme de la santé au Vénézuela: La transition Chávez (1999-2001)* [Unpublished doctoral dissertation]. Université de Montréal, Montréal, Canada.

Nápoles-Springer, A. M., Ortíz, C., O'Brien, H., & Díaz-Méndez, M. (2009). Developing a culturally competent peer support intervention for Spanish-speaking Latinas with breast cancer. *Journal of Immigrant and Minority Health*, *11*(4), 268–280. doi:10.1007/s10903-008-9128-4 PMID:18340533

National Breast and Cervical Cancer Early Detection Program. (2009, January 7). *Focus group research: Attitudes regarding mammography re-screening*. Retrieved from http://www.cdc.gov/dhdsp/CDCynergy_training/Content/activeinformation/resources/Mammography_Mod_Guide.pdf

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Nekhlyudov, L., & Braddock, C. H. III. (2009). An approach to enhance communication about screening mammography in primary care. *Journal of Women's Health*, *18*(9), 1403–1412. doi:10.1089/jwh.2008.1184 PMID:19708801

Nigenda, G., González-Robledo, M. C., González-Robledo, L. M., & Bejarano, R. M. (2011). *Disparidades al descubierto: Un studio comparado sobre políticas de cáncer de mama en América Latina*. Cuernavaca, México: Instituto Nacional de Salud Pública y American Cancer Society. Retrieved from http://ulaccam.org/wp-content/uploads/2012/03/reporte-Disaparidades-al-descubierto.pdf

OMNI Institute. (2004). *Essentials of planning and conducting focus groups*. Denver, CO: The OMNI Institute.

Otero-Sabogal, R., Stewart, S., Sabogal, F., Brown, B. A., & Pérez-Stable, E. J. (2003). Access and attitudinal factors related to breast and cervical cancer rescreening: Why are Latinas still underscreened? *Health Education & Behavior*, *30*(3), 337–359. doi:10.1177/1090198103030003008 PMID:19731500

Ott, J. J., Ullrich, A., & Miller, A. B. (2009). The importance of early symptom recognition in the context of early detection and cancer survival. *European Journal of Cancer*, *45*(16), 2743–2748. doi:10.1016/j.ejca.2009.08.009 PMID:19765977

Paulsel, M. L., McCroskey, J. C., & Richmond, V. P. (2006). Perceptions of healthcare professionals' credibility as a predictor of patients' satisfaction with their medical care and physician. *Communication Research Reports*, *23*(2), 69–76. doi:10.1080/08824090600668832

Plant, J. A. (2000). *Life in your hands: Understanding, preventing, and overcoming breast cancer*. New York: St. Martin's.

Provan, K. G., & Sebastian, J. G. (1998). Networks within networks: Service link overlap, organizational cliques, and network effectiveness. *Academy of Management Journal*, *41*(4), 453–463. doi:10.2307/257084

Robinson, J. D., Raup-Krieger, J. L., Burke, G., Weber, V., & Oesterling, B. (2008). The relative influence of patients' pre-visit global satisfaction on patients' post-visit satisfaction with physicians' communication. *Communication Research Reports*, *25*(1), 1–9. doi:10.1080/08824090701831735

Rogers, C., Thompson, K., & Robinson, S. (2002). Introducing a breast health strategy into schools. *Health Education*, *102*(3), 106–112. doi:10.1108/09654280210425994

Rohrer, J. E., Wilshusen, L., Adamson, S. C., & Merry, S. (2008). Patient-centredness, self-rated health, and patient empowerment: Should providers spend more time communicating with their patients? *Journal of Evaluation in Clinical Practice*, *14*(4), 548–551. doi:10.1111/j.1365-2753.2007.00914.x PMID:18462285

Rutten, L. J. F., Squiers, L., & Hesse, B. (2006). Cancer-related information seeking: Hints from the 2003 Health Information National Trends Survey (HINTS). *Journal of Health Communication*, *11*(S1), 147–156. doi:10.1080/10810730600637574 PMID:16641080

Shaw, B. R., McTavish, F., Hawkins, R., Gustafson, D. H., & Pingree, S. (2000). Experiences of women with breast cancer: Exchanging social support over the CHESS computer network. *Journal of Health Communication*, *5*(2), 135–159. doi:10.1080/108107300406866 PMID:11010346

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Slaughter, P., Pinfold, P., Flintoft, V., Gort, E., Thiel, E., Blackstien-Hirsch, P., & Williams, J. I. et al. (1999). *Focus groups in health services research at the institute for clinical evaluative sciences*. Toronto: Institute for Clinical Evaluative Sciences.

Small, S. A. (1995). Action-oriented research: Models and methods. *Journal of Marriage and the Family*, *57*(4), 941–955. doi:10.2307/353414

Small, S. A., & Uttal, L. (2005). Action-oriented research: Strategies for engaged scholarship. *Journal of Marriage and the Family*, *67*(4), 936–948. doi:10.1111/j.1741-3737.2005.00185.x

Sociedad Venezolana de Mastología. (2010). La Sociedad Venezolana de Mastología en tiempo de dificultades. *Carta Mastológica: Organo Divulgatorio de la Sociedad Venezolana de Mastología, 1*(28), 1-8. Retrieved from http://svmastologia.org/portal/images/stories/Cartas_Mastologicas/CM201007.pdf

Stang, I., & Mittlemark, M. B. (2010). Intervention to enhance empowerment in breast cancer self-help groups. *Nursing Inquiry*, *17*(1), 45–57. doi:10.1111/j.1440-1800.2009.00465.x PMID:20137030

Step, M. M., Rose, J. H., Albert, J. M., Cheruvu, V. K., & Siminoff, L. A. (2009). Modelling patient-centered communication: Oncologist relational communication and patient communication involvement in breast cancer adjuvant therapy decision-making. *Patient Education and Counseling*, *77*(3), 369–378. doi:10.1016/j.pec.2009.09.010 PMID:19811883

Street, R. L. (2003). Communicating in medical encounters: An ecological perspective. In T. L. Thompson, A. M. Dorsey, K. I. Miller, & R. Parrot (Eds.), *Handbook of health communication* (pp. 63–89). London: Lawrence Erlbaum Associates.

Street, R. L. Jr, Makoul, G., Arora, N. K., & Epstein, R. M. (2009). How does communication heal? Pathways linking clinician-patient communication to health outcomes. *Patient Education and Counseling*, *74*(3), 295–301. doi:10.1016/j.pec.2008.11.015 PMID:19150199

Swenson, S. L., Buell, S., Zettler, P., White, M., Ruston, D. C., & Lo, B. (2004). Patient-centered communication: Do patients really prefer it? *Journal of General Internal Medicine*, *19*(11), 1069–1079. doi:10.1111/j.1525-1497.2004.30384.x PMID:15566435

Talosig-Garcia, M., & Davis, S. W. (2005). Information-seeking behavior of minority breast cancer patients: An exploratory study. *Journal of Health Communication*, *10*(7), 53–64. doi:10.1080/10810730500263638 PMID:16377600

Thompson, S. (2009). Do you know why it's not just hair to me? The importance of eliciting breast-cancer survivors' explanatory models throughout the cancer continuum. *Journal of Community Health*, *2*(2), 148–158. doi:10.1179/cih.2009.2.2.148

Venetis, M. K., Robinson, J. D., Turkiewicz, K. L., & Allen, M. (2009). An evidence base for patient-centered cancer care: A meta-analysis of studies of observed communication between cancer specialists and their patients. *Patient Education and Counseling*, *77*(3), 379–383. doi:10.1016/j.pec.2009.09.015 PMID:19836920

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Vogel, B. A., Leonhart, R., & Helmes, A. W. (2009). Communication matters: The impact of communication and participation in decision making on breast cancer patients' depression and quality of life. *Patient Education and Counseling*, *77*(3), 391–397. doi:10.1016/j.pec.2009.09.005 PMID:19796910

Volkman, J. E., & Silk, K. J. (2008). Adolescent females and their mothers: Examining perceptions of the environment and breast cancer. *Journal of Health Psychology*, *13*(8), 1180–1189. doi:10.1177/1359105308095972 PMID:18987091

Warren, A., Londoño, G. E., Wessel, L. A., & Warren, R. D. (2006). Breaking down barriers to breast and cervical cancer screening: A university-based prevention program for Latinas. *Journal of Health Care for the Poor and Underserved*, *17*(3), 512–521. doi:10.1353/hpu.2006.0114 PMID:16960319

Washington, P. K., Burke, N. J., Galen, J., Guerra, C., & Pasick, R. J. (2009). Adult daughters' influence on mothers' health related decision making: An expansion of the subjective norms construct. *Health Education & Behavior*, *36*(1), 129S–144S. doi:10.1177/1090198109338904 PMID:19805795

Witt, J. (2008). The effect of information in the utilization of preventive health-care strategies: An application to breast cancer. *Health Economics*, *17*(6), 721–731. doi:10.1002/hec.1294 PMID:17990287

Wong, N.,, & King, T. (2008). The cultural construction of risk understandings through illness narratives. *The Journal of Consumer Research*, *34*(5), 579–594. doi:10.1086/520078

Zoller, H. M. (2005). Health activism: Communication theory and action for social change. *Communication Theory*, *15*(4), 341–364. doi:10.1111/j.1468-2885.2005.tb00339.x

ENDNOTES

[1] This action-oriented research project was supported by research grants from the Canadian Institutes of Health Research. It also received additional funding from private-sector donors in Venezuela such as Avon Venezuela, Roche, and Novartis.

[2] This four-year (2009-2012) action-oriented research project was developed by Professor Isaac Nahon-Serfaty and Professor Mahmoud Eid from the University of Ottawa (Canada) in partnership with the Venezuelan Breast Cancer NGO, SenosAyuda, with support from Professors Andrés Cañizales, Alfredo Poggi, José Rafael Briceño, Acianela Montes de Oca, and Néstor Garrido from the School of Social Communication of the Universidad Católica Andrés Bello (Caracas, Venezuela), and the Colegio Nacional de Periodistas, the association representing professional journalists in Venezuela.

[3] Patient-centered care is an approach to healthcare and treatment that places emphasis on patients' emotions, values, ethics, personal preferences, and needs to empower patients in their healing journey (Venetis et al., 2009).

[4] In a healthcare context, empowerment refers to feelings of control, confidence, and self-efficacy in one's health and healing, allowing patients to manage their bodies more effectively and make better decisions for themselves (Rohrer et al., 2008).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

[5] Patients' reluctance to communicate effectively with a physician is not necessarily a failing of the patient, but of the improper communication skills and training of the physician (e.g., Hanoch, 2004; Huijer & van Leeuwen, 2000).

[6] Medical language can often seem foreign and complex to patients, with jargon clouding their understanding of diagnoses and treatment options (Lee & Hawkins, 2010).

[7] Latin American women understand social support as providing women with informational guidance, emotional support, empathy and caring, and tangible help such as bringing women to appointments and picking up medications (Gonzalez, Gallardo & Bastani, 2005).

[8] A total of 121 people participated in the four workshops (June 5-6, 2009; June 11-12, 2010; May 20-21, 2011; and February 24-25, 2012). They were 50 activists representing different NGOs and including survivors and patients, 23 physicians, 28 journalists, 9 scholars, 5 public officials, and 6 representatives from the private sector.

This work was previously published in the International Journal of Civic Engagement and Social Change (IJCESC), 2(1); edited by Susheel Chhabra, pages 43-64, copyright year 2015 by IGI Publishing (an imprint of IGI Global).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Section 4
Emerging Trends

This section highlights research potential within the field of oncology while exploring uncharted areas of study for the advancement of the discipline. With 9 chapters the readers are introduced to future research directions and topical suggestions for continued debate, centering on the new venues and forums for discussion. In all, this text will serve as a vital resource to practitioners and academics interested in the best practices and applications of the burgeoning field of oncology. The book concludes with a look ahead into the future of the oncology field.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 30
Molecular Docking at a Glance

Maryam Hamzeh-Mivehroud
Tabriz University of Medical Sciences, Iran

Babak Sokouti
Tabriz University of Medical Sciences, Iran

Siavoush Dastmalchi
Tabriz University of Medical Sciences, Iran

ABSTRACT

The current chapter introduces different aspects of molecular docking technique in order to give an overview to the readers about the topics which will be dealt with throughout this volume. Like many other fields of science, molecular docking studies has experienced a lagging period of slow and steady increase in terms of acquiring attention of scientific community as well as its frequency of application, followed by a pronounced era of exponential expansion in theory, methodology, areas of application and performance due to developments in related technologies such as computational resources and theoretical as well as experimental biophysical methods. In the following sections the evolution of molecular docking will be reviewed and its different components including methods, search algorithms, scoring functions, validation of the methods, and area of applications plus few case studies will be touched briefly.

INTRODUCTION

Molecules are seldom found in isolation and all over the chemical universe they are involved in intermolecular interactions in homogenous or heterogeneous systems observed in almost all of the materials around us. The biological systems are no exceptions to this rule and the life with no question is dependent on these interactions. One of the important types of the interactions takes place between biological macromolecules known as receptors and the corresponding ligands. Understanding the molecular modes of such interactions is one of the major goals in comprehending the mechanism of actions of drug substances and drug design and discovery pipeline. Designing bioactive compounds with a controlled interaction profile toward a target of interest with minimum level of adverse effects is of great importance. In this context, molecular docking as a valuable tool in structure based drug design provides a framework to

DOI: 10.4018/978-1-5225-0549-5.ch030

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

engineer such molecules by visualizing structural requirements at the atomic level. Such useful information can significantly facilitate the rational drug design processes. The current chapter will introduce the molecular docking technique and its importance in computer-aided drug design. Moreover, different types of molecular docking methods, search algorithms as well as scoring functions will be presented briefly. Different examples of molecular docking applications will be also highlighted.

BACKGROUND

The inevitable need for developing new drugs originates from their inadequate efficiency and toxicity in the body. The history of drug discovery dates back to thousands of years ago wherein the ancient civilization used to employ herbal medicine as remedies for treatment of diseases. In spite of historical role of the serendipity factor throughout the traditional drug discovery, it is not regarded as a major component in modern drug discovery (circulation 2009). Drug development from inception to market requires an average time of 10-15 years and a cost of approximately a billion dollars (Campbell, Lamb, & Joseph-McCarthy, 2014; Ou-Yang et al., 2012). Being time consuming, costly, and risky, this process combines variety of tools to work more rationally in the process of introducing new drugs to the market. Late twentieth and early twenty- first century faced the substantial changes in drug design and discovery processes. The philosophy behind the rational drug design lies in logical reasoning before synthesizing any therapeutic agent for the biological evaluation. It provides a framework for prioritizing of potential lead compounds as early as possible, based on rational guidelines by employing the computational and experimental methods (Hung & Chen, 2014; Veselovsky & Ivanov, 2003). In this context, computer-aided drug design (CADD) is of great importance in shortening the development process in terms of cost and timeline. The CADD is mainly divided into two categories; ligand-based drug design (LBDD) and structure-based drug design (SBDD). In the case of LBDD, there is no information about the target of interest, and the chemical/biological features of a set of ligands are analyzed. Quantative-structure activity relationship (QSAR) techniques and pharmacophore-based design are the typical examples of LBDD strategy. If the 3D structural information of the target of interest is available, SBDD can also be utilized. In this strategy, the main focus is the design of novel ligands guided by investigating the ligand-target mutual interactions (Gani, 2007; Hung & Chen, 2014; Wilson & Lill, 2011). The starting point of SBDD approach is to identify the active site of the target molecule based on the known 3D structure of the bound natural ligand and /or drug molecules, mostly via X-ray crystallography or NMR experiments (Campbell et al., 2014; Mandal, Moudgil, & Mandal, 2009).The availability of the geometry for the target binding site may provide useful information for designing novel ligands or screening ligands from a large collection of chemical compounds based on structural and chemical complementarity features between target and ligand molecules (Gani, 2007; Mandal et al., 2009; Veselovsky & Ivanov, 2003).The candidates for synthesis and further biological assays are chosen based on the predicted binding mode and affinity as well as synthetic feasibility (Campbell et al., 2014). The most routinely used techniques in the area of SBDD are but not limited to molecular modeling, molecular dynamics, and docking techniques (Wilson & Lill, 2011).

The molecular docking is one the crucial components of the *in silico* studies which is heavily used in modern drug design and discovery efforts. Current molecular docking techniques have passed the infancy

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

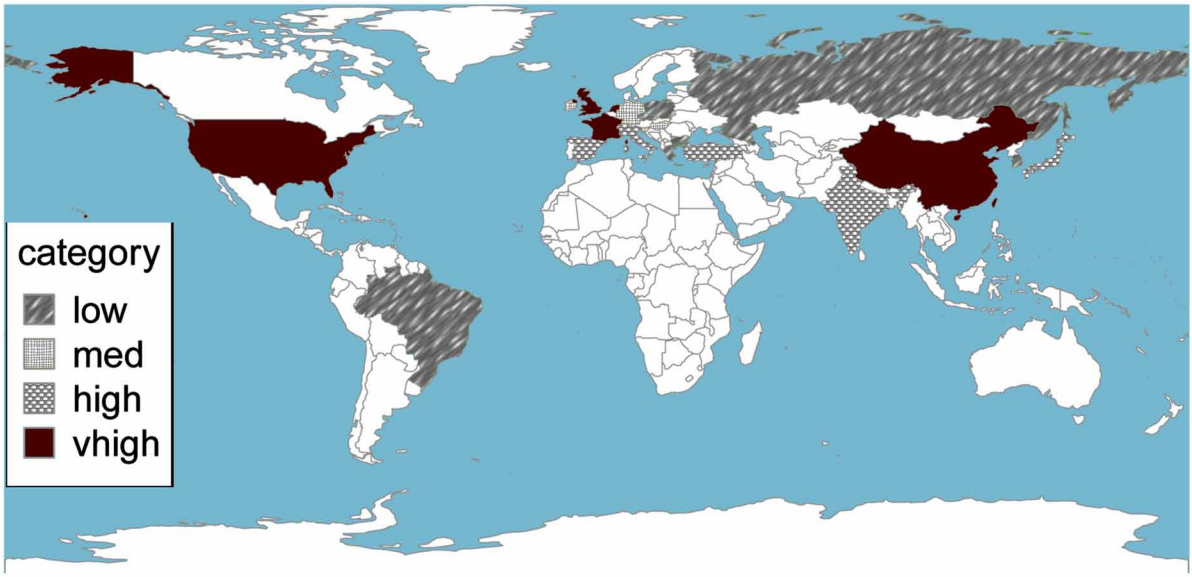
period and taking into account the tremendous advances in terms of novel computational algorithms as well as high performance computing hardware have led to its maturity to the state that is considered as an essential tool in rational drug design. In this method, the strength of intermolecular association is tried to be accurately predicted (Purisima & Hogues, 2012). A systematically driven statistics on the developed docking web servers all over the world and the fundamental principles of the molecular docking are described in the following sections.

WORLDWIDE USE OF MOLECULAR DOCKING

The numbers of docking web servers as a measure of popularity of docking calculations have been progressively increasing during the past decades. Figure 1 illustrates the distribution of the docking web servers based on main citation (original paper) all over the world during the period 2002-2015 which have also been listed in Table 1. In this list, docking programs are not included. Three top ranked cited web servers according to the country of origin are United States, United Kingdom and Netherlands, respectively.

To further appreciate the extent of penetration of docking studies in the scientific community, a simple query of PubMed for publications containing the terms “docking” and “ligand” was performed. The results shown in Figure 2 demonstrate that the number of docking publications from 1986 to 2014 has been increasing exponentially and it is estimated to reach about 700 by the end of 2015. This is just a naïve estimate and the exact statistics needs thorough inspection, however, it may provide a feeling for the widespread use of docking studies.

Figure 1. Docking Web servers based on citation



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 1. A comprehensive list of existing molecular docking webservers and related websites

No.	Name	Type	Website Address
1	NPDock(Tuszynska, Magnus, Jonak, Dawson, & Bujnicki, 2015)	protein-nucleic acid	genesilico.pl/NPDock
2	GalaxyPepDock(H. Lee, Heo, Lee, & Seok, 2015)	protein-peptide	galaxy.seoklab.org/cgi-bin/submit.cgi?type=PEPDOCK
3	CABS-dock(Kurcinski, Jamroz, Blaszczyk, Kolinski, & Kmiecik, 2015)	protein-peptide	biocomp.chem.uw.edu.pl/CABSdock
4	DOCKSCORE(Malhotra, Mathew, & Sowdhamini, 2015)	protein-protein	caps.ncbs.res.in/dockscore
5	pyDockSAXS(Jimenez-Garcia, Pons, Svergun, Bernado, & Fernandez-Recio, 2015)	protein-protein	life.bsc.es/pid/pydocksaxs
6	webSDA(X. Yu et al., 2015)	protein/nucleic acid-protein/ nucleic acid	mcm.h-its.org/webSDA
7	MTiOpenScreen(Labbe et al., 2015)	protein–ligand	bioserv.rpbs.univ-paris-diderot.fr/services/MTiOpenScreen
8	PSOVina(Ng, Fong, & Siu, 2015)	protein–ligand	cbbio.cis.umac.mo
9	14-3-3-Pred(Madeira et al., 2015)	phosphopeptides-protein	www.compbio.dundee.ac.uk/1433pred
10	AIDA(D. Xu, Jaroszewski, Li, & Godzik, 2014)	protein-protein	ffas.burnham.org/AIDA
11	ATTRACT(de Vries, Schindler, Chauvot de Beauchene, & Zacharias, 2015)	protein-protein	www.attract.ph.tum.de/services/ATTRACT/attract.html
12	CCharpPPI(Moal, Dapkunas, & Fernandez-Recio, 2015)	protein-protein	life.bsc.es/pid/ccharppi
13	CONSRANK(Chermak et al., 2015)	protein-protein/ nucleic acid	https://www.molnac.unisa.it/BioTools/consrank
14	SEABED(Fenollosa et al., 2015)	protein–ligand	www.bsc.es/SEABED
15	DOCKovalent(London et al., 2014)	protein–ligand	covalent.docking.org
16	CHARMMing(Pevzner, Frugier, Schalk, Caflisch, & Woodcock, 2014)	protein–ligand	www.charmming.org/charmming
17	ClusPro(Mottarella et al., 2014)	protein-Heparin	cluspro.bu.edu
18	PEP-SiteFinder(Saladin et al., 2014)	protein-peptide	bioserv.rpbs.univ-paris-diderot.fr/PEP-SiteFinder
19	PLIC(Anand, Nagarajan, Mukherjee, & Chandra, 2014)	protein-ligand	proline.biochem.iisc.ernet.in/PLIC
20	GalaxySite(Heo, Shin, Lee, & Seok, 2014)	protein-ligand	galaxy.seoklab.org/cgi-bin/submit.cgi?type=SITE
21	DOCK/PIERR(Viswanath, Ravikant, & Elber, 2014)	protein-protein	clsb.ices.utexas.edu/web/dock.html
22	SwarmDock(Torchala & Bates, 2014; Torchala, Moal, Chaleil, Fernandez-Recio, & Bates, 2013)	protein-protein	bmm.cancerresearchuk.org/~SwarmDock
23	SPOT-Seq-RNA(Y. Yang, Zhao, Wang, & Zhou, 2014)	protein-RNA	sparks-lab.org/yueyang/server/SPOT-Struct-RNA
24	DINC(Dhanik, McMurray, & Kavraki, 2013)	protein-ligand	dinc.kavrakilab.org
25	ZDOCK (B. G. Pierce et al., 2014)	protein-protein	zdock.umassmed.edu

continued on following page

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 1. Continued

No.	Name	Type	Website Address
26	Istar(H. Li, Leung, Ballester, & Wong, 2014)	protein-ligand	istar.cse.cuhk.edu.hk/idock
27	eSimDock(Brylinski, 2013)	protein-ligand	www.brylinski.org/esimdock
28	LigandRNA(Philips, Milanowska, Lach, & Bujnicki, 2013)	RNA-ligand	ligandrna.genesilico.pl
29	DockAFM(Chaves & Pellequer, 2013)	protein-protein/ DNA	biodev.cea.fr/dockafm
30	GOMoDo(Sandal et al., 2013)	GPCR protein-ligand	molsim.sci.univr.it/gomodo
31	DockRank(Xue, Jordan, El-Manzalawy, Dobbs, & Honavar, 2014)	protein-protein	einstein.cs.iastate.edu/index.html
32	Score-MI(Jardin, Stefani, Eberhardt, Huber, & Sticht, 2013)	protein-protein	score-mi.biochem.uni-erlangen.de
33	proABC(Olimpieri, Chailyan, Tramontano, & Marcatili, 2013)	protein-ligand	circe.med.uniroma1.it/proABC
34	ROSIE(Lyskov et al., 2013)	protein-ligand	rosie.rosettacommons.org
35	CovalentDock Cloud(Ouyang, Zhou, Ge, Li, & Kwoh, 2013)	protein-ligand	docking.sce.ntu.edu.sg
36	pyDockWEB(Jimenez-Garcia, Pons, & Fernandez-Recio, 2013)	protein-protein	life.bsc.es/servlet/pydock/home
37	EpiDOCK(Atanasova, Patronov, Dimitrov, Flower, & Doytchinova, 2013)	MHC class II-ligand	epidock.ddg-pharmfac.net/EpiDockPage.aspx
38	Sanjeevini(Jayaram et al., 2012)	protein/DNA-ligand	www.scfbio-iitd.res.in/sanjeevini/sanjeevini.jsp
39	FINDSITE(comb)(H. Zhou & Skolnick, 2013)	protein-ligand	cssb.biology.gatech.edu/FINDSITE-COMB
40	ClusPro protein(Brenke et al., 2012; Comeau, Gatchell, Vajda, & Camacho, 2004a)	protein-protein	cluspro.bu.edu/home.php
41	Quantum.Ligand.Dock(Kantardjiev, 2012)	protein-ligand	87.116.85.141/LigandDock.html
42	idTarget(Wang, Chu, Chen, & Lin, 2012)	protein-ligand	idtarget.rcas.sinica.edu.tw
43	Pose & Rank(Fan et al., 2011)	protein-ligand	modbase.compbio.ucsf.edu/poseandrank
44	BSP-SLIM(H. S. Lee & Zhang, 2012)	protein-ligand	zhanglab.ccmb.med.umich.edu/BSP-SLIM
45	AADS(Singh, Biswas, & Jayaram, 2011)	protein-ligand	www.scfbio-iitd.res.in/dock/ActiveSite_new.jsp
46	ProBiS(Konc, Konc, Penca, & Janezic, 2011)	protein-protein	probis.cmm.ki.si
47	KBDOCK(Ghoorah, Devignes, Smail-Tabbone, & Ritchie, 2011)	protein-protein	kbdock.loria.fr
48	GPU.proton.DOCK(Kantardjiev, 2011)	protein-protein docking	gpudock.orgchm.bas.bg
49	iScreen(Tsai, Chang, & Chen, 2011)	protein-ligand docking & virtual screening	iScreen.cmu.edu.tw
50	SDOCK(Zhang & Lai, 2011)	protein-protein docking	mdl.ipc.pku.edu.cn

continued on following page

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 1. Continued

No.	Name	Type	Website Address
51	Albumin binding prediction(Zsila et al., 2011)	human serum albumin (HSA) interactions	albumin.althotas.com
52	DEPTH(Tan, Varadarajan, & Madhusudhan, 2011)	depth-based predictions of binding sites	mspc.bii.a-star.edu.sg/depth
53	PRUNE and PROBE(Mitra & Pal, 2011)	protein-protein docking	pallab.serc.iisc.ernet.in/prune & pallab.serc.iisc.ernet.in/probe
54	CPORT(de Vries & Bonvin, 2011)	protein-protein docking	haddock.chem.uu.nl/services/CPORT
55	Erebus(Shirvanyants, Alexandrova, & Dokholyan, 2011)	reliably locate rigid binding scaffolds by searching PDB	erebus.dokhlab.org
56	TPACM4(Mukherjee, Patra, Barua, & Jayaram, 2011)	modeling interactions of small molecules with biomolecular target	www.scfbio-iitd.res.in/software/drugdesign/charge.jsp
57	FILTREST3D (Gajda, Tuszynska, Kaczor, Bakulina, & Bujnicki, 2010)	scoring and ranking of models	filtrest3d.genesilico.pl
58	BSR(Brylinski, Lee, Zhou, & Skolnick, 2011)	Binding Site Refinement	cssb.biology.gatech.edu/bsr
59	IBIS(Shoemaker et al., 2012)	protein-protein, protein-small molecule, protein nucleic acids and protein-ion interaction	www.ncbi.nlm.nih.gov/Structure/ibis/ibis.cgi
60	Opal(Ren, Williams, Clementi, Krishnan, & Li, 2010)	Include AutoDock and AutoDockVina	ws.nbcr.ne
61	POLYVIEW-MM(Porollo & Meller, 2010)	Automated mapping of ligand binding sites	polyview.cchmc.org/conform.html
62	RosettaBackrub(Lauck, Smith, Friedland, Humphris, & Kortemme, 2010)	flexible backbone protein structure modeling and design	kortemmelab.ucsf.edu/backrub
63	HexServer(Macindoe, Mavridis, Venkatraman, Devignes, & Ritchie, 2010)	exhaustive GPU-accelerated FFT-based rigid-body docking predictions	hexserver.loria.fr
64	ConPlex (Choi et al., 2010)	analyze protein complex structures based on evolutionary conservation of protein-protein interface residues	sbi.postech.ac.kr/ConPlex
65	HADDOCK(de Vries, van Dijk, & Bonvin, 2010)	protein-protein, protein-nucleic acids and protein-ligand complexes	haddock.science.uu.nl/services/HADDOCK2.2

continued on following page

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 1. Continued

No.	Name	Type	Website Address
66	PharmMapper(X. Liu et al., 2010)	potential drug target identification using pharmacophore mapping approach	59.78.96.61/pharmmapper
67	AuPosSOM (Bouvier, Evrard-Todeschi, Girault, & Bertho, 2010)	Automatic clustering of docking poses in virtual screening process	www.aupossom.com
68	POCASA (J. Yu, Zhou, Tanaka, & Yao, 2010)	predict binding sites by detecting pockets and cavities of proteins with a rolling spher	altair.sci.hokudai.ac.jp/g6/Research/POCASA_e.html
69	DOCK Blaster(Irwin et al., 2009)	Free Virtual Screening Server	blaster.docking.org
70	LIGCRYST(Sondergaard, Garrett, Carstensen, Pollastri, & Nielsen, 2009)	detecting crystal contacts in protein-ligand complexes	enzyme.ucd.ie/LIGCRYS
71	FRODOCK(Garzon et al., 2009)	fast rotational protein-protein docking	sbg.cib.csic.es/Software/FRODOCK
72	CT(Bayden, Fornabaio, Scarsdale, & Kellogg, 2009)	computational titration algorithm for preparing models for virtual screening and molecular docking	hinttools.isbdd.vcu.edu/CT
73	RosettaAntibody(Sircar, Kim, & Gray, 2009)	rational judgment in choosing the best model for docking	antibody.graylab.jhu.edu
74	@TOME-2(J. L. Pons & Labesse, 2009)	web server for comparative modeling of protein-ligand complexes	abcis.cbs.cnrs.fr/AT2
75	MRC(Rueda, Bottegoni, & Abagyan, 2009)	Generation of receptor conformation	abagyan.scripps.edu/MRC
76	SLITHER(P. H. Lee, Kuo, Chu, Liu, & Lin, 2009)	determining adjacent binding poses	bioinfo.mc.ntu.edu.tw/slither
77	DDNA 2.0(B. Xu, Yang, Liang, & Zhou, 2009)	protein-DNA docking	sparks.informatics.iupui.edu
78	RosettaDock (Lyskov & Gray, 2008)	protein-protein docking	rosettadock.graylab.jhu.edu
79	3D-Garden(Lesk & Sternberg, 2008)	protein-protein docking	www.sbg.bio.ic.ac.uk/3dgarden
80	ClusPro (Comeau et al., 2004a; Comeau, Gatchell, Vajda, & Camacho, 2004b)	protein-protein docking	cluspro.bu.edu/login.php

continued on following page

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 1. Continued

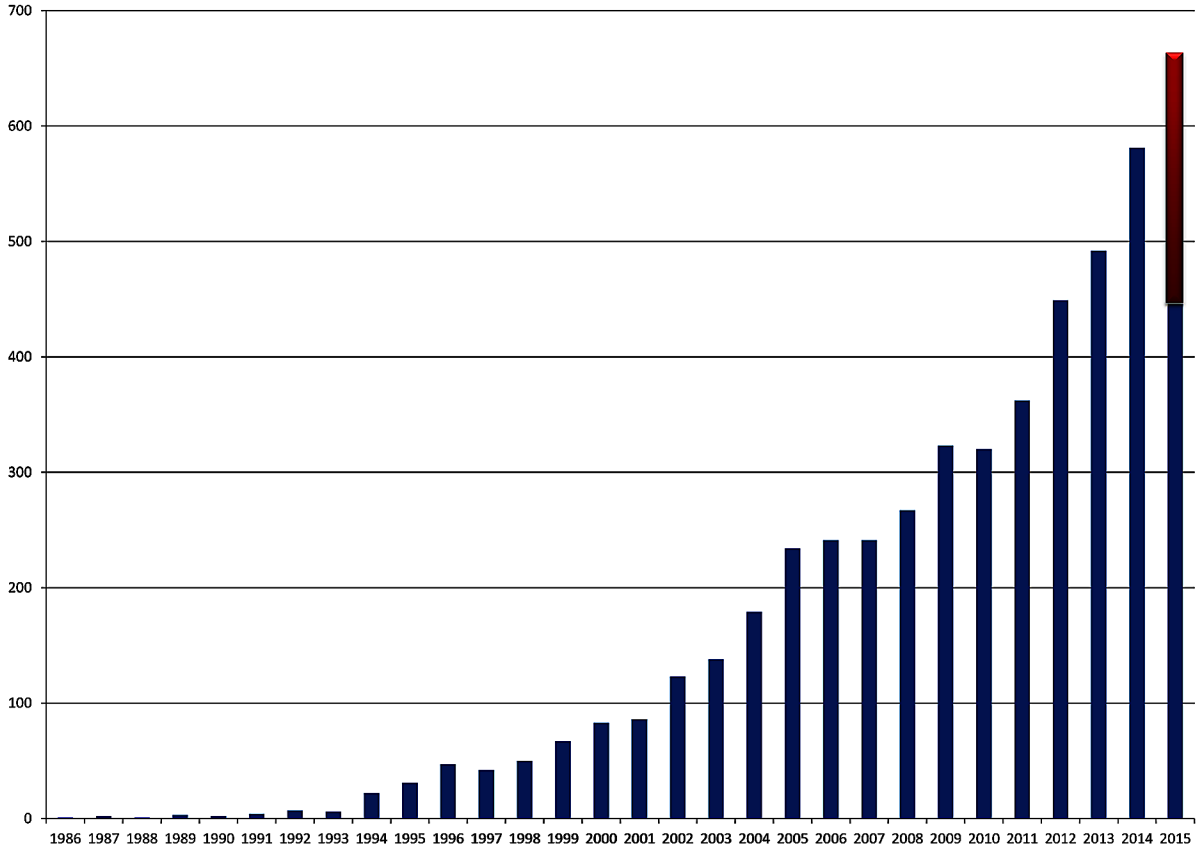
No.	Name	Type	Website Address
81	SKE-DOCK(Terashi et al., 2007)	protein-protein docking	www.pharm.kitasato-u.ac.jp/bmd/files/ske_dock.html
82	HingeProt(Emekli, Schneidman-Duhovny, Wolfson, Nussinov, & Haliloglu, 2008)	automated prediction of hinges in protein structures	bioinfo3d.cs.tau.ac.il/HingeProt
83	ROGER(Bernauer, Aze, Janin, & Poupon, 2007)	protein-protein docking scoring function	genomics.eu.org/spip/-Bioinformatics-tools
84	HADDOCK version 2.0(van Dijk & Bonvin, 2006)	Solvated docking	www.nmr.chem.uu.nl/haddock
85	KinDOCK (Martin, Catherinot, & Labesse, 2006)	analysis of ATP-binding sites of protein kinases	abcis.cbs.cnrs.fr/kindock
86	GRAMM-X(Tovchigrechko & Vakser, 2006)	protein-protein docking	vakser.bioinformatics.ku.edu/resources/gramm/grammx
87	TarFisDock(H. Li et al., 2006)	identifying drug targets with docking approach	www.dddc.ac.cn/tarfisdock
88	Scoring server(Camacho, Ma, & Champ, 2006)	scoring server for protein-protein interactions	structure.pitt.edu
89	Refining server(Lindahl & Delarue, 2005)	Refinement of docked protein-ligand and protein-DNA structures	lorentz.immstr.pasteur.fr
90	MEDock (Chang, Oyang, & Lin, 2005)	fficient prediction of ligand binding sites	bioinfo.mc.ntu.edu.tw/medock
91	Q-SiteFinder and Pocket-Finder(Laurie & Jackson, 2005)	prediction of protein-ligand binding sites	www.bioinformatics.leeds.ac.uk/pocketfinder
92	PPI_PRED(Bradford & Westhead, 2005)	prediction of protein-protein binding sites	www.bioinformatics.leeds.ac.uk/ppi_pred
93	ODA(Fernandez-Recio, Totrov, Skorodumov, & Abagyan, 2005)	predicting protein-protein interaction sites	www.molsoft.com/oda

MOLECULAR DOCKING PROCESS

The therapeutical outcome observed for the drug molecules comes from a very fundamental phenomenon known as molecular recognition. Regulation of cellular signal transduction pathways and cellular processes are governed by variety of such recognition events (Sousa et al., 2013). These events occur at the atomic based level and form the foundation of ligand-receptor interactions. Binding a ligand to an specific target is determined by shape and feature complementarity factors (C. Sotriffer, Klebe, Stahl, & Bohm, 2003). In the case of pharmaceuticals, the interactions are mostly established through non-covalent

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 2. An illustration of molecular docking publications during 1986-2015 (the red highlighted part in the plot demonstrates the estimated additional publications by the end of 2015)



interactions such as intermolecular van der Waals, hydrogen bonds, and electrostatic interactions (Brooijmans & Kuntz, 2003). Comprehensive study of such mutual ligand-receptor recognition originated from interactions is of great importance in structure-based drug design. One of the tools providing beneficial information on these interactions is molecular docking-based experiments.

Molecular docking is a term used for prediction of location, conformation, and orientation (totally referred as pose) of a molecule to the binding site of a target biomacromolecule (Houston & Walkinshaw, 2013; Kitchen, Decornez, Furr, & Bajorath, 2004).The history of the first molecular docking experiments dates back to the mid 1970s when protein-protein docking was performed (Levinthal, Wodak, Kahn, & Dadivanian, 1975; Salemme, 1976). The early docking methods were based on the "lock-and-key" theory proposed by Emil Fischer in 1894 (Fischer, 1894; Nabuurs, Wagener, & de Vlieg, 2007). In this theory, the shape of the receptor active site and the corresponding ligand are exactly complementary and the ligand fits to the receptor like a key and lock. This theory was shifted to "induced-fit" concept in which the ligand induces the conformational changes to the receptor during the binding process (Gani, 2007).

Different purposes are followed in the molecular docking area. Those which are the most important include: i) predicting the ligand binding mode within the receptor active site; ii) identifying the novel ligands through virtual screening; and iii) predicting the compounds' binding affinities from a known active ligand (Kitchen et al., 2004; Kroemer, 2003; Leach, Shoichet, & Peishoff, 2006).

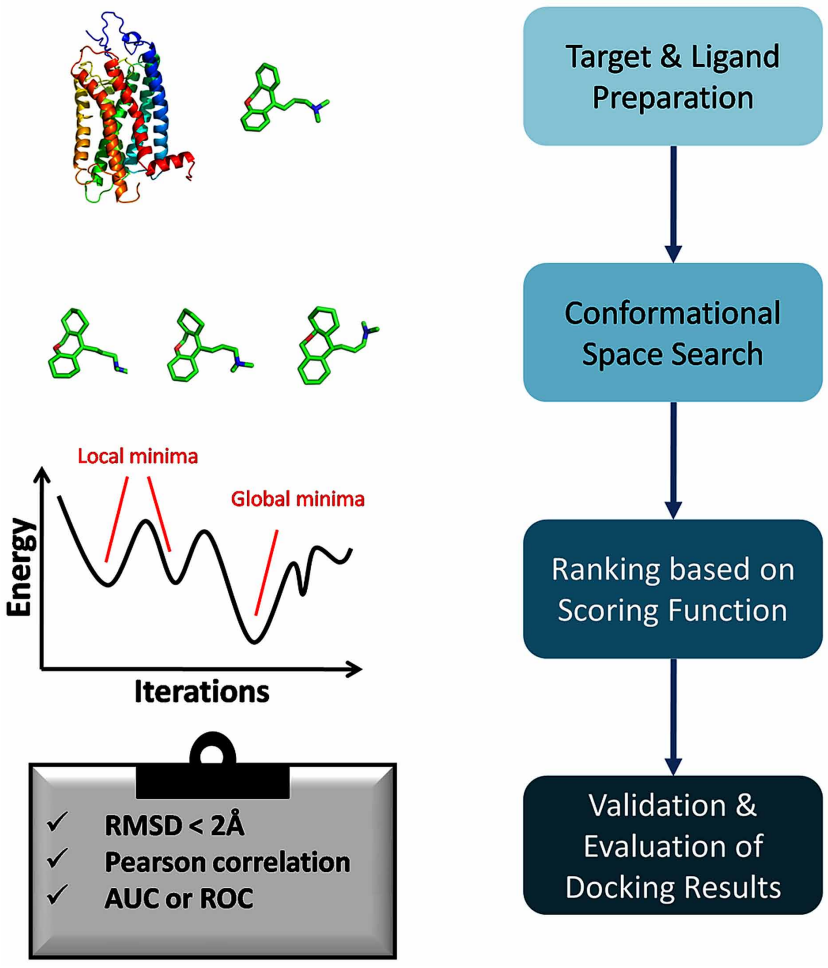
EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Any molecular docking experiment comprises several common key steps as listed below:

1. Target and ligand selection and representation,
2. Conformational space search,
3. Ranking of potential solution based on scoring functions, and
4. Validation and evaluation of docking results (G. M. Morris & Lim-Wilby, 2008).

Figure 3 shows a typical docking workflow. The following sections explain these steps with some more details.

Figure 3. The workflow of molecular docking process



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

MOLECULAR REPRESENTATION FOR MOLECULAR DOCKING

The two essential interplaying parts in any molecular docking are receptor and ligand molecules. The most widely used approaches for representation of the receptor relies on atomic, surface, and grid representations (Kitchen et al., 2004; C. Sotriffer et al., 2003). Atomic representation is generally used when the ranking is based on potential energy function. In some cases, surface based representation is employed where the rigid body approximation is considered. The grid representation is also applied based on potential energy grids in which the energy contribution of the receptor on the grid points (mainly electrostatic and van der Waals potentials) is calculated, then a map is generated for each atom or interaction types in order to provide the essential information for scoring during the ligand docking process (Goodford, 1985; C. Sotriffer et al., 2003).

In order to do a molecular docking experiment, the 3D-structure of the receptor which is used in the calculation is either an experimental (solved 3D structures of proteins deposited in the protein data bank) or a model structure predicted based on different prediction methodologies such as "comparative(homology and threading) modeling". In comparative method, the modeling is based on either a direct similarity between the sequences of protein of interest and that of a known structure called "template" whereas in threading or fold recognition method, the compatibility of the sequence of target protein with the structure of known folds (templates) in database is assessed (Kroemer, 2003). Three dimensional structure of protein can then be used for defining the binding site. If the structure of target along with its ligand is known (when the ligand co-crystallized with the target), the binding site is determined with certainty; otherwise, the binding site needs to be determined in several ways such as using tools to deduce protein function from structural homology or approaches derived from the inspection of physicochemical and geometrical characteristics of the receptor surface for locating the binding site. For example, cavities in the target protein, specially the biggest one, seem to be the most probable places to be defined as binding site. The catalytic site of the protein which is surrounded by the residues involved in catalysis can also be regarded as the binding site particularly if the inhibitors' docking is sought (G. M. Morris & Lim-Wilby, 2008; Veselovsky & Ivanov, 2003).

The other essential component of the molecular docking is ligand. Two strategies have been used for ligand handling: fragment-based methods and whole-molecule approaches. In both methods the binding site is considered to be known. In fragment-based method, a library of fragments is employed and the way the docking is completed proceeds by one of the following methods. In the first method, fragments are docked in a separate fashion followed by reconnection of them through searching the most favorable combination in terms of minimum energy and allowed degrees of freedom. In another method, the ligand is expanded increasingly during docking process using fragments referred to "anchor and grow". In whole-molecule approach, the entire molecule is docked as a whole (Brooijmans & Kuntz, 2003; G. M. Morris & Lim-Wilby, 2008; C. Sotriffer et al., 2003)(8, 9, 16). Different levels of approximation in terms of flexibility are used for whole molecule approaches which will be discussed in the following sections.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

DOCKING CLASSIFICATION METHODS

Flexibility

In a molecular docking experiment, the flexibility is defined as six degrees of translational (along *x*-, *y*-, and *z*- axes) and rotational freedom. Moreover, each of ligand and receptor has its conformational degrees of freedom depending on torsion angles of each rotatable bond (Gani, 2007; Leach et al., 2006; G. M. Morris & Lim-Wilby, 2008). Since the search for all conformational space is a difficult task requiring spending much more time and computational cost, an approximation for reducing the dimensionality of the space is essential (Sousa, Fernandes, & Ramos, 2006).One of the most common classifications of molecular docking algorithms is based on extent of molecular flexibility. From the flexibility point of view, three levels of approximation in docking methods exist including: *i)* rigid docking *ii)* semi-flexible docking and *iii)* flexible docking. Earliest and simplest approaches of molecular docking were based on rigid-body approximation. In rigid docking the structural conformation of ligand and receptor are not changed during the docking procedure and both of them regarded as rigid solid bodies. In second approach i.e. semi-flexible docking, which is extensively used, the receptor is assumed as rigid structure while the conformational change will be occurred for the ligand molecule. In the case of flexible docking, conformational degrees of freedom along with translational and rotational degrees of freedom are also considered for both of ligand and receptor (Hung & Chen, 2014; Leach et al., 2006; Sousa et al., 2013). Accuracy and speed in molecular docking are two main limiting factors. Rigid-body treatment of two components (i.e. ligand and receptor) in molecular docking is faster than condition where the flexibility is included in the docking methodology (G. M. Morris & Lim-Wilby, 2008). This is due to smaller size of search space. Although high speed of a docking experiment is an advantage, the effect of different conformations of both ligand and receptor in unbound and bound states cannot be ignored. Upon binding of the ligand to the receptor (usually is a protein), significant conformational change such as rearrangement of the side chains and motion of loops and domains in the receptor structure can occur (see Koshland's induced-fit theory)(Lill, 2011). As a result, for an ideal system the flexibility of both ligand and receptor should be taken into account since there are several degrees of freedom in a real biological system. For this purpose, developing the algorithms capable of implementing these considerations is of great value. Soft-receptor modeling, side-chain flexibility, Monte Carlo (MC), molecular dynamics (MD), as well as ensemble grids docking are different categories employed in the existing methodologies (Huang & Zou, 2010a; Sousa et al., 2006). However, the receptor flexibility is still in its infancy and only limited motion for the receptor is considered in the currently used algorithms because of computing time and coverage of all conformational space. In order to enter the flexibility of the ligand, several algorithms have been developed consist of: systematic methods, random or stochastic methods, and simulation methods (Sousa et al., 2006). These methods and above mention methods will be discussed in the following sections.

Interacting Partners

Since the molecular bases of a biological response comes from the interactions formed between biological macromolecules such as enzymes, antibodies, nucleic acids and ligands which span a wide range in terms of size and molecular nature, the other classification of docking depends on considering these features of the interacting partners. Therefore, different types of docking from molecular perspective are:

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

small molecule (e.g. drug or drug-like molecules)-protein, protein-protein, DNA/RNA-protein, and small molecule-DNA/RNA interactions (Ricci & Netz, 2009; Sousa et al., 2006; R. D. Taylor, Jewsbury, & Essex, 2002; Tuffery & Derreumaux, 2012). In most of the cases, small molecules constitute one of the participating elements in the docking experiments (Velec, Gohlke, & Klebe, 2005; Wang et al., 2012). High flexibility of small ligands makes them to be treated in small ligand-oriented docking schemes as flexible or semi-flexible. In addition, for small ligand-protein docking, the complementary contact interface is much smaller and less discriminating compared to the protein-protein docking. However, predicting three-dimensional binding pocket of the ligandis the main goal of small ligand-protein docking (Tuffery & Derreumaux, 2012). Protein-protein interactions are the source of many cellular functions and important in understanding the mechanism of signaling pathways. The most important issue in the field of protein-protein docking is presence of huge number of degrees of freedom making it impractical due to inability searching the entire conformational space. Because of this, rigid body handling of protein-protein docking is usually applied in many of the algorithms. Despite this fact, several attempts have been performed to tackle this problem (Baxter, Murray, Clark, Westhead, & Eldridge, 1998; Jerabek, Florian, Stiborova, & Martinek, 2014). Nucleic acids are another important type of targets for therapeutics like anti-bacterial, chemotherapeutic, and antiviral agents, and also is being recognized by many biological molecules such as proteins, the interaction of nucleic acids with protein or small molecules is of particular importance (Ricci & Netz, 2009). They are similar to proteins in terms of binding specificity (supplied by specific nucleic acid binding motifs) and conformational changes upon binding, but due to high charge density and unique helical geometry characteristics are distinct from proteins. Moreover, they exhibit no certain binding site and are highly solvent exposed. Studies showed that preferable binding site of protein to DNA is in its major groove whereas small ligands interact with minor groove of the DNA double helical structures (Ricci & Netz, 2009).

A BRIEF INTRODUCTION TO MOLECULAR DOCKING METHODS

Despite the various types of molecular docking protocols (i.e., protein-protein, protein-ligand, protein-RNA, protein-DNA, and etc) in which any mutual parts can be rigid or flexible, the main workflow includes searching for the correct pose with the most stable state (i.e., global minimum energy), and then evaluating the predicted conformations by relevant scoring functions for differentiating between true and all the other binding modes (Kitchen et al., 2004; Sousa et al., 2006). The computational cost of docking for placing two molecules near each other is exponentially high considering various degrees of freedoms including translational and rotational poses (Halperin, Ma, Wolfson, & Nussinov, 2002; Huang, 2014). This will be even more challenging when the flexibility of both molecules is considered which necessitates a need for exploring a huge conformational search space. As a rigorous method, one may exhaustively generate and explore all the possible binding modes considering their three translational and three rotational freedoms of the targets of interest, however this approach will be unfeasible considering the computational power of recently designed hardware. For an example provided by Taylor and his colleague who conducted a review on the protein-small molecule docking methods, by considering the 10 degree increments of angles and 6×10^{14} conformations, it will take about 2×10^{6} years for a computer with the capability of processing 10 conformations per second to complete the computations (R. D. Taylor et al., 2002). As a result, an essence is felt for providing a balance between the computational costs and the explored search space by developing efficient search algorithms as well as scoring functions.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Search Strategies

The search algorithms in the docking process can be mainly divided into two approaches. The first approach is to search for the potential solution candidates in the full solution space namely "exhaustive/systematic search" and the second approach is to search the solution space gradually or randomly using the intelligent or evolutionary methodologies (e.g., Monte Carlo (MC), genetic algorithm (GA), simulated annealing (SA), and Tabu search) which in turn only a part of the solution space is explored to reduce the computational costs (Halperin et al., 2002; Zsoldos, Reid, Simon, Sadjad, & Johnson, 2006). However, over a last decade, another field has been also emerged which is supposed to optimize the docking process further by considering it as a multi-objective problem (MOP) instead of only a one-objective problem (Nicolaou & Brown, 2013) due to the fact that drug discovery should be regarded as a challenging MOP with various pharmaceutical objectives (Schneider & Fechner, 2005). Considering the exhaustive search methodology, that samples all of the conformations with some/many unfavorable states, the optimization algorithm may not have a successful result in addition to the high cost of computational operations. For this purpose, in the early years of docking approaches, the algorithms were benefitted from the rigid bodies of both target and receptors to reduce the dimensionality (R. D. Taylor et al., 2002) to only six degrees of translational and rotational freedom which had its own shortcomings (R. D. Taylor et al., 2002). However, recently two subsets of direct search strategies in Cartesian space namely "local shape feature matching" and "Post-docking approach" have been discussed by Huang (Huang, 2014).

To reduce the highly computational cost of the exhaustive search (i.e., a standard search method) on exploring and evaluating the high number of binding orientations of complexes, *(i)* fast/spherical Fourier transform correlation and *(ii)* direct search in Cartesian space approaches have been introduced (Huang, 2014). In 1992, for speeding up the search process, a group of researchers proposed a mapping strategy from translational space to imaginary space for searching the target space using fast Fourier transform (FFT) calculations (Katchalski-Katzir et al., 1992). Some of the FFT-based docking programs with common principle behind their search strategies using a grid mapping structure are ZDOCK (Chen, Li, & Weng, 2003; Chen, Tong, Mintseris, Li, & Weng, 2003; Hwang, Vreven, Pierce, Hung, & Weng, 2010; B. Pierce, Tong, & Weng, 2005; B. G. Pierce, Hourai, & Weng, 2011; B. G. Pierce et al., 2014; Vreven, Pierce, Hwang, & Weng, 2009, 2013; Wiehe et al., 2005; Wiehe et al., 2007; Yi, Qiu, Cao, Wu, & Li, 2008; Zhang, Liu, & Zhou, 2005), SDOCK (Zhang & Lai, 2011), Cell-Dock (C. Pons et al., 2012), F2Dock (Bajaj, Chowdhury, & Siddavanahalli, 2011), PIPER (Kozakov, Brenke, Comeau, & Vajda, 2006), PIEDOCK (Viswanath et al., 2014), F(2)Dock2.0 protocol used for managing multiple docking jobs and visualizing the docked poses and interfaces (Chowdhury et al., 2013), and commercial ICM-Pro from Molsoft (Fernandez-Recio, Totrov, & Abagyan, 2003; Mendez, Leplae, De Maria, & Wodak, 2003; Mendez, Leplae, Lensink, & Wodak, 2005) from which ZDOCK has received much attention in docking research studies in recent decades. Although, the FFT-based search approach is applied over the 3D translational space in the above mentioned programs, however, it can also be applied over either the 3D rotational space (e.g., FRODOCK (Garzon et al., 2009) or both of the 3D translational and rotational spaces (e.g., HEX (Macindoe et al., 2010; Ritchie & Kemp, 2000; Ritchie, Kozakov, & Vajda, 2008; Ritchie & Venkatraman, 2010)). Furthermore, the ZDOCK has taken the advantage of the 3D convolutional library for accelerating the computational FFT calculations (B. G. Pierce et al., 2011) and recently, the graphics processing unit architecture and programming have been emerged for speeding up the parallel computations of FFT operations (Kantardjiev, 2011, 2012; Ritchie & Venkatraman, 2010; Ritchie, Venkatraman, & Mavridis, 2010; Shimoda, Suzuki, Ohue, Ishida, & Akiyama, 2015). Direct

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

search in Cartesian space approach is known to be simpler but not more efficient than FFT-based methods in terms of computational costs while using the Boolean operators and heuristic rules with protein flexibility, however both of them use the Cartesian grid space as their mapping tool but the direct search looks for the shape matches between two proteins in the grid-based mapped system (Huang, 2014).

In the shape matching search approach, with its computational efficient advantage, the goal is to quickly position the target protein represented by molecular shapes (e.g., Connolly surface) in the binding site in its good shape complimentarily considering the six degrees of the freedom(Huang, 2014; Huang & Zou, 2010a). And, if only a match is found and a binding orientation is generated, the transformation matrix of six degrees of freedom will be calculated which makes it more different from the exhaustive search algorithms (Huang, 2014). Examples of shape matching docking programs include DOCK (Kuntz, Blaney, Oatley, Langridge, & Ferrin, 1982), Surflex (Jain, 2003, 2007; Miteva, Lee, Montes, & Villoutreix, 2005; Spitzer & Jain, 2012), MDock(Huang & Zou, 2011), GAPDOCK (Gardiner, Willett, & Artymiuk, 2003), and FLOG (Miller, Kearsley, Underwood, & Sheridan, 1994).

Systematic search algorithms using exhaustive search approach is regarded as one of the time consuming and infeasible methods, therefore, fragmentation and database methods can also be included from which fragmentation based techniques are the most commonly used and continued to be a hot topic in docking studies (Sousa et al., 2006; Yuriev, Holien, & Ramsland, 2015). The fragment-based approach uses an incremental or stepwise algorithm to dock various molecular fragments covalently into the active site or divide them into two parts namely "core fragment" and "side chains" in which the core fragments (i.e., rigid part) are docked first and then the side chains (i.e., flexible parts) are added incrementally (Kitchen et al., 2004). Examples of fragment-based approaches include LUDI (Bohm, 1992a, 1992b), FlexX(Schellhammer & Rarey, 2004), DOCK (Kuntz et al., 1982), ADAM (Mizutani, Tomioka, & Itai, 1994), Hammerhead (Welch, Ruppert, & Jain, 1996), and eHiTs (Zsoldos et al., 2006). Database methods are used to minimize the computational costs of exhaustive search methodologies by taking advantage of pre-generated libraries of ligand conformations, for example, FLOG program produces 25 conformations for each molecule using the distance geometry (Miller et al., 1994; Sousa et al., 2006).

Unlike the previously mentioned search approaches, the randomized search algorithms which search at atomic level can be used either with or without molecular representations (i.e., grid or shape) to speed up the search computational process (Huang, 2014). The randomized search algorithms work based on a probability/fitness function that is optimized by using several heuristic algorithms such as Monte Carlo (MC) based on Metropolis criterion using the Boltzmann probability function (e.g., ICM (Abagyan, Totrov, & Kuznetsov, 2004), MCDOCK (M. Liu & Wang, 1999)), genetic algorithm (GA) (e.g.,GOLD(Jones, Willett, & Glen, 1995; Jones, Willett, Glen, Leach, & Taylor, 1997), AutoDock (Garrett M. Morris et al., 1999), FIPSDOCK (Y. Liu et al., 2013), MoDOCK (Gu, Yang, Kang, Wu, & Wang, 2015)), particle swarm ((Garcia-Godoy, Lopez-Camacho, Garcia-Nieto, & Aldana-Montes, 2015; Meier, Pippel, Brandt, Sippl, & Baldauf, 2010; Namasivayam & Gunther, 2007), Tabu search methods (e.g., PRO_LEADS (Baxter et al., 1998; Murray, Baxter, & Frenkel, 1999)) (Sousa et al., 2006), and combination of other algorithms such as GA or geometry search with Tabu search (Hou, Wang, Chen, & Xu, 1999; Tondel, Anderssen, & Drablos, 2006). The advantage of Tabu search is that the revisiting conformations are recorded as a list based on their RMSD values and only new conformations from the search space will be accepted and analyzed (Sousa et al., 2006).

Molecular dynamics (MD) is a computer simulation technique for modeling the complicated molecular systems at the atomic level which is categorized in the simulation search methods. In this approach, the solutions to Newton's equations of motions are calculated for finding the global minimum energy of a

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

docked complex (R. D. Taylor et al., 2002). However, most of the times, the solutions are trapped in a local minimum of the energy surface (Brooijmans & Kuntz, 2003) and cannot cover the high-energy conformational barriers which counts as a main disadvantage for the MD technique (R. D. Taylor et al., 2002). As a result, the quality of the solutions will be dependent on the starting point of the molecular conformations. In some studies, the authors have overcome the shortcomings of MD simulation methods by utilizing a jumping method for not being trapped in local minima, or simulate parts of the docked system at different temperatures (Di Nola, Roccatano, & Berendsen, 1994) and taken advantage of MD performance in comparison to other methods (Jerabek et al., 2014; Pak & Wang, 2000). Another category of the simulation search methods is the energy minimization methodology which is a less widely applicable technique alone at its own (i.e., gradient, conjugate gradient, second derivatives, and least square methods (Sousa et al., 2006)), in contrast to the MD simulation method (Kitchen et al., 2004) and reaches the local minima solutions if not used with other evolutionary algorithms. However, there are some other new approaches that employ the well-known Monte Carlo method along with the energy minimization method for achieving the global minimum of the energy surface such as GlamDock (Tietze & Apostolakis, 2007) or using the Bezier splines based approach (Trosset & Scheraga, 1998). Most of the common docking algorithms such as QXP (McMartin & Bohacek, 1997), DARWIN (J. S. Taylor & Burnett, 2000), CRDOCK (Cortes Cabrera et al., 2012), and ADAM (Mizutani et al., 1994) use various types of energy minimization methods for their process.

The post-docking approach is a two stage protocol for refining the docking results and hence, it may not be directly categorized in the search algorithms. In this approach, by using a docking program all possible conformations are generated and in the next step, the top binding conformations of the first step are re-ranked based on other features such as biological information using relevant scoring methods (Huang, 2014). Some of the post-docking approaches which have also been assessed in Critical Assessment of PRedicted Interaction (CAPRI) experiments are RDOCK (L. Li, Chen, & Weng, 2003), PyDock (Cheng, Blundell, & Fernandez-Recio, 2007; C. Pons, Solernou, Perez-Cano, Grosdidier, & Fernandez-Recio, 2010), SmoothDock(Camacho & Gatchell, 2003), and MultiDock (Jackson, Gabb, & Sternberg, 1998).

Scoring Functions and Their Evaluations

During the above mentioned search algorithm process, a broad unmanageable range of binding modes of the conformations will be generated, however, generating of these conformations will not be sufficient and the correct conformation should be also recognizable (Sousa et al., 2006). For this purpose, an accurate and speedy tool for predicting the binding free energies (e.g., within 7-10 kJ/mol(Bissantz, Folkers, & Rognan, 2000)) known as “scoring function” is needed to discriminate between the correct (i.e., true positive) and the remaining false positive solutions (Halperin et al., 2002).On the other hand, a rigorous scoring function for calculating the energy values of all conformations is neither computationally cost effective nor with low complexity. So, by considering the speed and accuracy properties of scoring functions which is a main bottleneck in molecular docking studies (Kitchen et al., 2004), some pre-definitions, assumptions, and simplifications should be considered in order to reduce the complexity order in developing the scoring functions (Sousa et al., 2006). The various types of algorithms utilized in scoring functions can be divided into three main categories including force field, empirical, and knowledge-based scoring functions (Huang & Zou, 2010a).

Among various types of developed force field scoring function approaches (e.g., Tripos(Kramer, Rarey, & Lengauer, 1999) and AMBER (Weiner, Kollman, Nguyen, & Case, 1986) force fields), the

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

summation value of two types of energies namely the "receptor-ligand interaction energy" and the "internal ligand energy" is denoted as a standard molecular mechanics force field (Kitchen et al., 2004). In most of the scoring approaches, the internal protein energy is simply ignored for reducing the calculation costs. However, the absence of solvation and entropic terms as well as not accurate treatment of long-range interactions involved in binding counts as one of the main shortcomings of standard force field scoring functions (Sousa et al., 2006). To overcome this limitation, recently, torsional entropy and protein-ligand hydrogen-bonding terms have been implemented in programs such as G-Score (Kramer et al., 1999), GOLD (Verdonk, Cole, Hartshorn, Murray, & Taylor, 2003) and AutoDock (G. M. Morris, Goodsell, Huey, & Olson, 1996; G. M. Morris et al., 2013; G. M. Morris et al., 2009), to increase the possibility of molecular recognition.

The second main types of scoring functions are the empirical functions which predict the affinity binding score based on a linear combination/or linear regression of multiple weighted energy terms (e.g., van der Waals, electrostatics, desolvation, hydrogen binding, entropy, and hydrophobicity (Huang, Grinter, & Zou, 2010). However, the performance of the empirical scoring functions solely depends on the quality of the training set (i.e., the training set includes thousands of diverse protein-ligand complexes with known 3D structures) which fits the affinity binding scores to their corresponding features of the training set and the issues related many energy terms (Huang et al., 2010). Due to simple and linear structure of these functions, they are more efficient than force field scoring functions in terms of computational costs, speed, and accuracy (Huang et al., 2010; Vajda, Hall, & Kozakov, 2013). Some of the examples which are developed based on the empirical scoring functions are LUDI (Bohm, 1992b), MedusaScore(Yin, Biedermannova, Vondrasek, & Dokholyan, 2008), AIScore (Raub, Steffen, Kamper, & Marian, 2008), and LigScore (Krammer, Kirchhoff, Jiang, Venkatachalam, & Waldman, 2005).

The rationale behind the knowledge-based scoring functions is to define the potential of mean force (Zsoldos et al., 2006) by means of the statistical mechanics in the physics field which has been satisfied by employing the inverse Boltzmann method (Huang et al., 2010; McQuarrie, 1976). In this technique, the pairwise energy potentials are extracted in terms of rules from the experimental structures at the atomic level as well as their occurrence frequencies of atom pairs to calculate the scores of ligand binding poses (Huang et al., 2010; Zsoldos et al., 2006). Some of the common examples are DrugScore developed based on the Cambridge Structural Database (Gohlke & Klebe, 2002; Nisius & Gohlke, 2012; C. A. Sotriffer, Gohlke, & Klebe, 2002; Velec et al., 2005), SMoG/SMoG2001 (DeWitte & Shakhnovich, 1996), ITScore/SE (Huang & Zou, 2010b), PMF (Muegge & Martin, 1999) as well as QSAR studies for scoring the protein-ligand interactions (Ballester & Mitchell, 2010) –that are developed based on the knowledge-based scoring functions.

Moreover, recently, the performance of scoring functions have been improved using software or hardware based techniques such as consensus (combination of multiple scoring functions such as GFScore (Betzi, Suhre, Chetrit, Guerlesquin, & Morelli, 2006), PLS-DA (Avram, Pacureanu, Seclaman, Bora, & Kurunczi, 2011), VoteDock (Plewczynski, Lazniewski, von Grotthuss, Rychlewski, & Ginalski, 2011), CPORT (de Vries & Bonvin, 2011)), clustering and entropy-based (e.g., CyClus (Omori & Kitao, 2013)), machine learning based scoring functions (Ballester & Mitchell, 2010; Hsin, Ghosh, & Kitano, 2013; Khamis, Gomaa, & Ahmed, 2015; Kinnings et al., 2011), parallel computing on graphics processing units (GPUs) (e.g., Quantum.Ligand.Dock (Kantardjiev, 2012), GPU.proton.DOCK (Kantardjiev, 2011), and other GPU accelerated studies (Ritchie et al., 2010; Shimoda et al., 2015; Sukhwani & Herbordt, 2010; Wu, Hong, Takeda, & Guo, 2012)), and multi objective optimization programming (MOOP) (e.g., MoDock (Gu et al., 2015) and other MOOP approaches (H. Li et al., 2009; Nicolaou & Brown, 2013)).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

For evaluating the performance of the scoring functions, there are three types of measurement criteria including *(i)* binding mode identification, *(ii)* binding affinity prediction, and *(iii)* virtual database screening (Huang et al., 2010). In the first criterion which is widely used in the (Huang, 2014) molecular docking studies, it is important to discriminate between the binding modes and decoys based on the predicted RMSD values (i.e., if the RMSD value is less than 2 Å, the prediction is regarded as successful and otherwise it is not). The second criterion is to measure the ability of the scoring function in predicting the binding affinity by the means of Pearson correlation between the predicted and actual experimental scores. The third approach is virtual screening to identify the potential hits in a large database. Moreover, the area under curve (AUC) and the area under the receiver operating characteristics (ROC) are also other common scoring function testing approaches used in virtual screening of drug design. If the performance of a scoring function is successfully validated by all of the three above mentioned techniques, the obtained results are highly reliable, however, in practice, it is very rare to find a scoring function to satisfy all of the three measurement criteria (Huang et al., 2010).

MOLECULAR DOCKING APPLICATIONS

Molecular docking is one of the foundations of the modern structure-based drug discovery. This technique tries to envisage the inter-molecular interactions in a virtual manner and has attracted enormous attention in drug discovery pipeline. Molecular docking can be utilized for many different purposes, but all may be distilled down to an *in silico* effort for theprediction of ligand binding mode. Having the binding mode predicted enables using the results for different aims. The binding affinity, as a very fundamental measure in ligand-target interaction, is one of the main properties, which can be dealt with using the predicted poses derived from the docking studies. For example, in a study by Politi and her coworkers, the binding affinity for a series of renin inhibitors was predicted through molecular docking study (Politi, Durdagi, Moutevelis-Minakakis, Kokotos, & Mavromoustakos, 2010). Yang et al. investigated the prediction of the binding affinities for small-molecule Smac (the second mitochondria-derived activator of caspase) mimetics toward the target using this technique (C. Y. Yang, Sun, Chen, Nikolovska-Coleska, & Wang, 2009).These are just few examples and obviously presenting a comprehensive list of such studies is impossible in this introductory chapter.

From qualitative point of view, there are numerous examples in which different conformations at different poses on the target molecule are generated and the energetically favored pose or poses are selected using a scoring function to identify the most probable interactions between the ligand and its target molecule (Brooijmans & Kuntz, 2003; Hamzeh-Mivehroud, Mahmoudpour, & Dastmalchi, 2012; Hamzeh-Mivehroud, Rahmani, Feizi, Dastmalchi, & Rashidi, 2014; Hamzeh-Mivehroud, Rahmani, Rashidi, Hosseinpour Feizi, & Dastmalchi, 2013).

Another important application of molecular docking is its significant contribution in virtual screening. Virtual screening is a multistep task in which molecular docking of a large data set of drug-like molecules (also called virtual chemical library) on drug targets is performed followed by post processing of the poses based on calculated binding affinities for selecting the top ranking ligands (Muegge & Enyedy, 2003; Schmidt, Bergner, & Schwede, 2014). This is performed to find molecules having *druglikeness* property capable of binding to a given target. If the candidate molecules (referred to as hits) or its modified relatives show good affinity to the target of interest in biochemical assays, they can be further evaluated as the potential "lead" compounds (Muegge & Enyedy, 2003). The lead optimization can be performed

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

through structural modification in order to improve its biological activity, as well as its other properties collectively known as ADMET[1]. There are many studies in which molecular docking has served as a tool in virtual screening (Bleicher, Bohm, Muller, & Alanine, 2003; Brylinski, 2013; Elokely & Doerksen, 2013; Houston & Walkinshaw, 2013; Pirhadi, Shiri, & Ghasemi, 2013), however, the followings are just a few examples. Recently, Wei et al. have used an optimized hierarchical multistage virtual screening method targeting HIV-1 protease. Six compounds were selected for *in vitro* HIV-1 protease inhibitory assay and among them two molecules showed inhibitory activities against HIV-1 protease, which may be suitable for structural modification in drug development process (Wei et al., 2015). A similar work was conducted to find novel compounds against the signal transducer and activator of transcription 3 (STAT3) leading to identification of STX-0119 as STAT3 dimerization inhibitor. Anti-tumor activity of this compound by *in vivo* evaluation showed promising result in cancer treatment and can enter the structure optimization process for improved potency and ADMET properties (Matsuno et al., 2010). In an attempt to identify inhibitors of influenza virus PA endonuclease, docking based virtual screening method was used. Among the selected fifteen compounds as new chemical entities two of them had IC_{50} values in micromolar range (Pala et al., 2015). For more comprehensive examples related to lead identification and optimization via molecular docking-based virtual screening see (Laurini et al., 2011; Nolan et al., 2011; Ou-Yang et al., 2012; Yakar & Akten, 2014).

Molecular docking experiments are also extensively utilized in 3D-QSAR analyses as a molecular alignment tool. In an investigation by Patel et al. inhibitory activity for benzimidazole derivatives as Hepatitis C Virus NS5B polymerase inhibitors were predicted using 3D-QSAR, where molecular docking was used for aligning the molecules (Patel, Patel, Kaushik-Basu, & Talele, 2008). In another study, the flexible docking of the nonnucleoside inhibitors of HIV-1 reverse transcriptase was performed in order to determine the active conformation of the studied ligands, and then the results were used in 3D-QSAR analysis (Z. Zhou & Madura, 2004). For more examples see (Harpsoe, Liljefors, & Balle, 2008; Tintori, Magnani, Schenone, & Botta, 2009; Tuccinardi et al., 2007; Vitorovic-Todorovic, Cvijetic, Juranic, & Drakulic, 2012).

The above mentioned case studies are few examples of molecular docking applications in structure-based drug design. However, as mentioned above, the citation is not comprehensive by any means.

FUTURE RESEARCH DIRECTIONS

Molecular docking, an on the move computational-experimental approach, is one of the most attention-received research fields in terms of quantity and the development of search algorithms as well as scoring functions. However, despite these advances there are still many challenges and limitations which need to be well addressed to improve the performance quality of the docking approaches in terms of accuracy and speed. To this end progress in the development of novel search strategies, scoring functions, and representation of protein flexibility as well as obtaining high accuracy with low computational costs by employing any software and hardware technologies, is needed simultaneously. Among these, the quality of the information in the search space is also a problem, as the more information does not always include the better information.

Among different types of search algorithms, the randomized search strategy may reserve the accuracy with high computational cost because the molecular representation is at the atomic level instead of grid based mapping approach which includes the complicated projection to grids and losing the accuracy of

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

scoring functions. Some of the solutions which show increasing role in improving the docking results include (Huang, 2014): i) integration of scoring functions with sampling/searching processes; ii) Using additional information about binding by restricting the sampling space and template-based docking (i.e., these information can also be added in FFT/SFT based search algorithms); and iii) Considering the flexibility of target proteins at either receptor or ligand sides.

Moreover, the transition from considering the rigid to flexible bodies whether it is only the receptor or the ligand or both of them, will add more computational costs on all of the docking processes. Regarding the improvement of scoring functions, the intelligent machine learning approaches form an active and open area in molecular docking experiments to accelerate the docking process with high accuracy which also one may speed them up by utilizing several parallel and GPU based hardware improvement approaches.

In conclusion, a balance between accuracy and speed should be considered while developing new and high performance algorithms both in exploring the search space and in optimizing the selections by scoring functions.

CONCLUSION

Drug discovery looks like "finding a needle in a haystack" and is considered to be a time-consuming, tricky and complicated process. For the pharmaceutical companies the time and financial expenses are the main important criteria in the drug discovery and development projects with high failure rate. In modern drug discovery and development pipeline both of biophysical and computational approaches are employed side by side for discovering the bioactive compounds. Digitization of the biological systems has profound effect on shortening the way to identify therapeutic agent by reducing the number of compounds needed to be synthesized and biologically evaluated. In this sense, computational approaches have shed new light on drug design and discovery process and tremendous advances have been achieved during the past decades. Molecular docking as an invaluable tool in structure-based drug design has become a first line technique in early phase of hit and lead identification since the number of biological structures (as targets) in data banks are growing. Availability of the three dimensional structure for the target molecule responsible for the observed biological event, increases the possibility and accelerates the design of molecules, which can bind to the active site of the target and modulate its activity. Despite the fact that great progress and promising accomplishment has been made in this area, undoubtedly, the molecular docking and other related *in silico* prediction techniques cannot replace the information gathered by experimental investigation. We are far from the perfect and ideal molecular docking method and the issues dealing with fully flexible molecular docking in terms of computational cost and accuracy are still challenging.

REFERENCES

Abagyan, R., Totrov, M., & Kuznetsov, D. (2004). ICM—A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *Journal of Computational Chemistry*, *15*(5), 488–506. doi:10.1002/jcc.540150503

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Anand, P., Nagarajan, D., Mukherjee, S., & Chandra, N. (2014). PLIC: Protein-ligand interaction clusters. *Database (Oxford)*, *2014*(0), bau029. doi:10.1093/database/bau029 PMID:24763918

Atanasova, M., Patronov, A., Dimitrov, I., Flower, D. R., & Doytchinova, I. (2013). EpiDOCK: A molecular docking-based tool for MHC class II binding prediction. *Protein Engineering, Design & Selection*, *26*(10), 631–634. doi:10.1093/protein/gzt018 PMID:23661105

Avram, S., Pacureanu, L. M., Seclaman, E., Bora, A., & Kurunczi, L. (2011). PLS-DA - Docking Optimized Combined Energetic Terms (PLSDA-DOCET) protocol: A brief evaluation. *Journal of Chemical Information and Modeling*, *51*(12), 3169–3179. doi:10.1021/ci2002268 PMID:22066983

Bajaj, C., Chowdhury, R., & Siddavanahalli, V. (2011). F2Dock: Fast Fourier protein-protein docking. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *8*(1), 45–58. doi:10.1109/TCBB.2009.57 PMID:21071796

Ballester, P. J., & Mitchell, J. B. (2010). A machine learning approach to predicting protein-ligand binding affinity with applications to molecular docking. *Bioinformatics (Oxford, England)*, *26*(9), 1169–1175. doi:10.1093/bioinformatics/btq112 PMID:20236947

Baxter, C. A., Murray, C. W., Clark, D. E., Westhead, D. R., & Eldridge, M. D. (1998). Flexible docking using Tabu search and an empirical estimate of binding affinity. *Proteins*, *33*(3), 367–382. doi:10.1002/(SICI)1097-0134(19981115)33:3<367::AID-PROT6>3.0.CO;2-W PMID:9829696

Bayden, A. S., Fornabaio, M., Scarsdale, J. N., & Kellogg, G. E. (2009). Web application for studying the free energy of binding and protonation states of protein-ligand complexes based on HINT. *Journal of Computer-Aided Molecular Design*, *23*(9), 621–632. doi:10.1007/s10822-009-9270-7 PMID:19554265

Bernauer, J., Aze, J., Janin, J., & Poupon, A. (2007). A new protein-protein docking scoring function based on interface residue properties. *Bioinformatics (Oxford, England)*, *23*(5), 555–562. doi:10.1093/bioinformatics/btl654 PMID:17237048

Betzi, S., Suhre, K., Chetrit, B., Guerlesquin, F., & Morelli, X. (2006). GFscore: A general nonlinear consensus scoring function for high-throughput docking. *Journal of Chemical Information and Modeling*, *46*(4), 1704–1712. doi:10.1021/ci0600758 PMID:16859302

Bissantz, C., Folkers, G., & Rognan, D. (2000). Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *Journal of Medicinal Chemistry*, *43*(25), 4759–4767. doi:10.1021/jm001044l PMID:11123984

Bleicher, K. H., Bohm, H. J., Muller, K., & Alanine, A. I. (2003). Hit and lead generation: Beyond high-throughput screening. *Nature Reviews. Drug Discovery*, *2*(5), 369–378. doi:10.1038/nrd1086 PMID:12750740

Bohm, H. J. (1992a). The computer program LUDI: A new method for the de novo design of enzyme inhibitors. *Journal of Computer-Aided Molecular Design*, *6*(1), 61–78. doi:10.1007/BF00124387 PMID:1583540

Bohm, H. J. (1992b). LUDI: Rule-based automatic design of new substituents for enzyme inhibitor leads. *Journal of Computer-Aided Molecular Design*, *6*(6), 593–606. doi:10.1007/BF00126217 PMID:1291628

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Bouvier, G., Evrard-Todeschi, N., Girault, J. P., & Bertho, G. (2010). Automatic clustering of docking poses in virtual screening process using self-organizing map. *Bioinformatics (Oxford, England)*, *26*(1), 53–60. doi:10.1093/bioinformatics/btp623 PMID:19910307

Bradford, J. R., & Westhead, D. R. (2005). Improved prediction of protein-protein binding sites using a support vector machines approach. *Bioinformatics (Oxford, England)*, *21*(8), 1487–1494. doi:10.1093/bioinformatics/bti242 PMID:15613384

Brenke, R., Hall, D. R., Chuang, G. Y., Comeau, S. R., Bohnuud, T., Beglov, D., & Kozakov, D. et al. (2012). Application of asymmetric statistical potentials to antibody-protein docking. *Bioinformatics (Oxford, England)*, *28*(20), 2608–2614. doi:10.1093/bioinformatics/bts493 PMID:23053206

Brooijmans, N., & Kuntz, I. D. (2003). Molecular recognition and docking algorithms. *Annual Review of Biophysics and Biomolecular Structure*, *32*(1), 335–373. doi:10.1146/annurev.biophys.32.110601.142532 PMID:12574069

Brylinski, M. (2013). Nonlinear scoring functions for similarity-based ligand docking and binding affinity prediction. *Journal of Chemical Information and Modeling*, *53*(11), 3097–3112. doi:10.1021/ci400510e PMID:24171431

Brylinski, M., Lee, S. Y., Zhou, H., & Skolnick, J. (2011). The utility of geometrical and chemical restraint information extracted from predicted ligand-binding sites in protein structure refinement. *Journal of Structural Biology*, *173*(3), 558–569. doi:10.1016/j.jsb.2010.09.009 PMID:20850544

Camacho, C. J., & Gatchell, D. W. (2003). Successful discrimination of protein interactions. *Proteins*, *52*(1), 92–97. doi:10.1002/prot.10394 PMID:12784373

Camacho, C. J., Ma, H., & Champ, P. C. (2006). Scoring a diverse set of high-quality docked conformations: A metascore based on electrostatic and desolvation interactions. *Proteins*, *63*(4), 868–877. doi:10.1002/prot.20932 PMID:16506242

Campbell, A. J., Lamb, M. L., & Joseph-McCarthy, D. (2014). Ensemble-based docking using biased molecular dynamics. *Journal of Chemical Information and Modeling*, *54*(7), 2127–2138. doi:10.1021/ci400729j PMID:24881672

Chang, D. T., Oyang, Y. J., & Lin, J. H. (2005). MEDock: a web server for efficient prediction of ligand binding sites based on a novel optimization algorithm. *Nucleic Acids Res, 33*, W233-238.

Chaves, R. C., & Pellequer, J. L. (2013). DockAFM: Benchmarking protein structures by docking under AFM topographs. *Bioinformatics (Oxford, England)*, *29*(24), 3230–3231. doi:10.1093/bioinformatics/btt561 PMID:24078683

Chen, R., Li, L., & Weng, Z. (2003). ZDOCK: An initial-stage protein-docking algorithm. *Proteins*, *52*(1), 80–87. doi:10.1002/prot.10389 PMID:12784371

Chen, R., Tong, W., Mintseris, J., Li, L., & Weng, Z. (2003). ZDOCK predictions for the CAPRI challenge. *Proteins*, *52*(1), 68–73. doi:10.1002/prot.10388 PMID:12784369

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Cheng, T. M., Blundell, T. L., & Fernandez-Recio, J. (2007). pyDock: Electrostatics and desolvation for effective scoring of rigid-body protein-protein docking. *Proteins*, *68*(2), 503–515. doi:10.1002/prot.21419 PMID:17444519

Chermak, E., Petta, A., Serra, L., Vangone, A., Scarano, V., Cavallo, L., & Oliva, R. (2015). CONSRANK: A server for the analysis, comparison and ranking of docking models based on inter-residue contacts. *Bioinformatics (Oxford, England)*, *31*(9), 1481–1483. doi:10.1093/bioinformatics/btu837 PMID:25535242

Choi, Y. S., Han, S. K., Kim, J., Yang, J. S., Jeon, J., Ryu, S. H., & Kim, S. (2010). ConPlex: a server for the evolutionary conservation analysis of protein complex structures. *Nucleic Acids Res, 38*, W450-456. doi: 10.1093/nar/gkq328

Chowdhury, R., Rasheed, M., Keidel, D., Moussalem, M., Olson, A., Sanner, M., & Bajaj, C. (2013). Protein-protein docking with F(2)Dock 2.0 and GB-rerank. *PLoS ONE*, *8*(3), e51307. doi:10.1371/journal.pone.0051307 PMID:23483883

Comeau, S. R., Gatchell, D. W., Vajda, S., & Camacho, C. J. (2004a). ClusPro: a fully automated algorithm for protein-protein docking. *Nucleic Acids Res, 32*, W96-99. doi: 10.1093/nar/gkh354

Comeau, S. R., Gatchell, D. W., Vajda, S., & Camacho, C. J. (2004b). ClusPro: An automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics (Oxford, England)*, *20*(1), 45–50. doi:10.1093/bioinformatics/btg371 PMID:14693807

Cortes Cabrera, A., Klett, J., Dos Santos, H. G., Perona, A., Gil-Redondo, R., Francis, S. M., & Morreale, A. et al. (2012). CRDOCK: An ultrafast multipurpose protein-ligand docking tool. *Journal of Chemical Information and Modeling*, *52*(8), 2300–2309. doi:10.1021/ci300194a PMID:22764680

de Vries, S. J., & Bonvin, A. M. (2011). CPORT: A consensus interface predictor and its performance in prediction-driven docking with HADDOCK. *PLoS ONE*, *6*(3), e17695. doi:10.1371/journal.pone.0017695 PMID:21464987

de Vries, S. J., Schindler, C. E., Chauvot de Beauchene, I., & Zacharias, M. (2015). A web interface for easy flexible protein-protein docking with ATTRACT. *Biophysical Journal*, *108*(3), 462–465. doi:10.1016/j.bpj.2014.12.015 PMID:25650913

de Vries, S. J., van Dijk, M., & Bonvin, A. M. (2010). The HADDOCK web server for data-driven biomolecular docking. *Nature Protocols*, *5*(5), 883–897. doi:10.1038/nprot.2010.32 PMID:20431534

DeWitte, R. S., & Shakhnovich, E. I. (1996). SMoG: De Novo Design Method Based on Simple, Fast, and Accurate Free Energy Estimates. 1. Methodology and Supporting Evidence. *Journal of the American Chemical Society*, *118*(47), 11733–11744. doi:10.1021/ja960751u

Dhanik, A., McMurray, J. S., & Kavraki, L. E. (2013). DINC: A new AutoDock-based protocol for docking large ligands. *BMC Structural Biology*, *13*(Suppl 1), S11. doi:10.1186/1472-6807-13-S1-S11 PMID:24564952

Di Nola, A., Roccatano, D., & Berendsen, H. J. (1994). Molecular dynamics simulation of the docking of substrates to proteins. *Proteins*, *19*(3), 174–182. doi:10.1002/prot.340190303 PMID:7937732

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Elokely, K. M., & Doerksen, R. J. (2013). Docking challenge: Protein sampling and molecular docking performance. *Journal of Chemical Information and Modeling*, *53*(8), 1934–1945. doi:10.1021/ci400040d PMID:23530568

Emekli, U., Schneidman-Duhovny, D., Wolfson, H. J., Nussinov, R., & Haliloglu, T. (2008). HingeProt: Automated prediction of hinges in protein structures. *Proteins*, *70*(4), 1219–1227. doi:10.1002/prot.21613 PMID:17847101

Fan, H., Schneidman-Duhovny, D., Irwin, J. J., Dong, G., Shoichet, B. K., & Sali, A. (2011). Statistical potential for modeling and ranking of protein-ligand interactions. *Journal of Chemical Information and Modeling*, *51*(12), 3078–3092. doi:10.1021/ci200377u PMID:22014038

Fenollosa, C., Oton, M., Andrio, P., Cortes, J., Orozco, M., & Goni, J. R. (2015). SEABED: Small molEcule activity scanner weB servicE baseD. *Bioinformatics (Oxford, England)*, *31*(5), 773–775. doi:10.1093/bioinformatics/btu709 PMID:25348211

Fernandez-Recio, J., Totrov, M., & Abagyan, R. (2003). ICM-DISCO docking by global energy optimization with fully flexible side-chains. *Proteins*, *52*(1), 113–117. doi:10.1002/prot.10383 PMID:12784376

Fernandez-Recio, J., Totrov, M., Skorodumov, C., & Abagyan, R. (2005). Optimal docking area: A new method for predicting protein-protein interaction sites. *Proteins*, *58*(1), 134–143. doi:10.1002/prot.20285 PMID:15495260

Fischer, E. (1894). Einfluss der Configuration auf die Wirkung der Enzyme. *Berichte der deutschen chemischen Gesellschaft, 27*(3), 2985-2993.

Gajda, M. J., Tuszynska, I., Kaczor, M., Bakulina, A. Y., & Bujnicki, J. M. (2010). FILTREST3D: Discrimination of structural models using restraints from experimental data. *Bioinformatics (Oxford, England)*, *26*(23), 2986–2987. doi:10.1093/bioinformatics/btq582 PMID:20956242

Gani, O. A. (2007). Signposts of docking and scoring in drug design. *Chemical Biology & Drug Design*, *70*(4), 360–365. doi:10.1111/j.1747-0285.2007.00571.x PMID:17937781

Garcia-Godoy, M. J., Lopez-Camacho, E., Garcia-Nieto, J., & Aldana-Montes, A. J. (2015). Solving molecular docking problems with multi-objective metaheuristics. *Molecules (Basel, Switzerland)*, *20*(6), 10154–10183. doi:10.3390/molecules200610154 PMID:26042856

Gardiner, E. J., Willett, P., & Artymiuk, P. J. (2003). GAPDOCK: A Genetic Algorithm Approach to Protein Docking in CAPRI round 1. *Proteins*, *52*(1), 10–14. doi:10.1002/prot.10386 PMID:12784360

Garzon, J. I., Lopez-Blanco, J. R., Pons, C., Kovacs, J., Abagyan, R., Fernandez-Recio, J., & Chacon, P. (2009). FRODOCK: A new approach for fast rotational protein-protein docking. *Bioinformatics (Oxford, England)*, *25*(19), 2544–2551. doi:10.1093/bioinformatics/btp447 PMID:19620099

Ghoorah, A. W., Devignes, M. D., Smail-Tabbone, M., & Ritchie, D. W. (2011). Spatial clustering of protein binding sites for template based protein docking. *Bioinformatics (Oxford, England)*, *27*(20), 2820–2827. doi:10.1093/bioinformatics/btr493 PMID:21873637

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Gohlke, H., & Klebe, G. (2002). DrugScore meets CoMFA: Adaptation of fields for molecular comparison (AFMoC) or how to tailor knowledge-based pair-potentials to a particular protein. *Journal of Medicinal Chemistry*, *45*(19), 4153–4170. doi:10.1021/jm020808p PMID:12213058

Goodford, P. J. (1985). A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *Journal of Medicinal Chemistry*, *28*(7), 849–857. doi:10.1021/jm00145a002 PMID:3892003

Gu, J., Yang, X., Kang, L., Wu, J., & Wang, X. (2015). MoDock: A multi-objective strategy improves the accuracy for molecular docking. *Algorithms for Molecular Biology; AMB*, *10*(1), 8. doi:10.1186/s13015-015-0034-8 PMID:25705248

Halperin, I., Ma, B., Wolfson, H., & Nussinov, R. (2002). Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins*, *47*(4), 409–443. doi:10.1002/prot.10115 PMID:12001221

Hamzeh-Mivehroud, M., Mahmoudpour, A., & Dastmalchi, S. (2012). Identification of new peptide ligands for epidermal growth factor receptor using phage display and computationally modeling their mode of binding. *Chemical Biology & Drug Design*, *79*(3), 246–259. doi:10.1111/j.1747-0285.2011.01282.x PMID:22136656

Hamzeh-Mivehroud, M., Rahmani, S., Feizi, M. A., Dastmalchi, S., & Rashidi, M. R. (2014). In vitro and in silico studies to explore structural features of flavonoids for aldehyde oxidase inhibition. *Archiv der Pharmazie*, *347*(10), 738–747. doi:10.1002/ardp.201400076 PMID:24956178

Hamzeh-Mivehroud, M., Rahmani, S., Rashidi, M. R., Hosseinpour Feizi, M. A., & Dastmalchi, S. (2013). Structure-based investigation of rat aldehyde oxidase inhibition by flavonoids. *Xenobiotica*, *43*(8), 661–670. doi:10.3109/00498254.2012.755228 PMID:23282065

Harpsoe, K., Liljefors, T., & Balle, T. (2008). Prediction of the binding mode of biarylpropylsulfonamide allosteric AMPA receptor modulators based on docking, GRID molecular interaction fields and 3D-QSAR analysis. *Journal of Molecular Graphics & Modelling*, *26*(5), 874–883. doi:10.1016/j.jmgm.2007.06.002 PMID:17644444

Heo, L., Shin, W. H., Lee, M. S., & Seok, C. (2014). GalaxySite: ligand-binding-site prediction by using molecular docking. *Nucleic Acids Res, 42*, W210-214.

Hou, T., Wang, J., Chen, L., & Xu, X. (1999). Automated docking of peptides and proteins by using a genetic algorithm combined with a tabu search. *Protein Engineering*, *12*(8), 639–648. doi:10.1093/protein/12.8.639 PMID:10469824

Houston, D. R., & Walkinshaw, M. D. (2013). Consensus docking: Improving the reliability of docking in a virtual screening context. *Journal of Chemical Information and Modeling*, *53*(2), 384–390. doi:10.1021/ci300399w PMID:23351099

Hsin, K. Y., Ghosh, S., & Kitano, H. (2013). Combining machine learning systems and multiple docking simulation packages to improve docking prediction reliability for network pharmacology. *PLoS ONE*, *8*(12), e83922. doi:10.1371/journal.pone.0083922 PMID:24391846

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Huang, S. Y. (2014). Search strategies and evaluation in protein-protein docking: Principles, advances and challenges. *Drug Discovery Today*, *19*(8), 1081–1096. doi:10.1016/j.drudis.2014.02.005 PMID:24594385

Huang, S. Y., Grinter, S. Z., & Zou, X. (2010). Scoring functions and their evaluation methods for protein-ligand docking: Recent advances and future directions. *Physical Chemistry Chemical Physics*, *12*(40), 12899–12908. doi:10.1039/c0cp00151a PMID:20730182

Huang, S. Y., & Zou, X. (2010a). Advances and challenges in protein-ligand docking. *International Journal of Molecular Sciences*, *11*(8), 3016–3034. doi:10.3390/ijms11083016 PMID:21152288

Huang, S. Y., & Zou, X. (2010b). Inclusion of solvation and entropy in the knowledge-based scoring function for protein-ligand interactions. *Journal of Chemical Information and Modeling*, *50*(2), 262–273. doi:10.1021/ci9002987 PMID:20088605

Huang, S. Y., & Zou, X. (2011). Construction and test of ligand decoy sets using MDock: Community structure-activity resource benchmarks for binding mode prediction. *Journal of Chemical Information and Modeling*, *51*(9), 2107–2114. doi:10.1021/ci200080g PMID:21755952

Hung, C. L., & Chen, C. C. (2014). Computational approaches for drug discovery. *Drug Development Research*, *75*(6), 412–418. doi:10.1002/ddr.21222 PMID:25195585

Hwang, H., Vreven, T., Pierce, B. G., Hung, J. H., & Weng, Z. (2010). Performance of ZDOCK and ZRANK in CAPRI rounds 13-19. *Proteins*, *78*(15), 3104–3110. doi:10.1002/prot.22764 PMID:20936681

Irwin, J. J., Shoichet, B. K., Mysinger, M. M., Huang, N., Colizzi, F., Wassam, P., & Cao, Y. (2009). Automated docking screens: A feasibility study. *Journal of Medicinal Chemistry*, *52*(18), 5712–5720. doi:10.1021/jm9006966 PMID:19719084

Jackson, R. M., Gabb, H. A., & Sternberg, M. J. (1998). Rapid refinement of protein interfaces incorporating solvation: Application to the docking problem. *Journal of Molecular Biology*, *276*(1), 265–285. doi:10.1006/jmbi.1997.1519 PMID:9514726

Jain, A. N. (2003). Surflex: Fully automatic flexible molecular docking using a molecular similarity-based search engine. *Journal of Medicinal Chemistry*, *46*(4), 499–511. doi:10.1021/jm020406h PMID:12570372

Jain, A. N. (2007). Surflex-Dock 2.1: Robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search. *Journal of Computer-Aided Molecular Design*, *21*(5), 281–306. doi:10.1007/s10822-007-9114-2 PMID:17387436

Jardin, C., Stefani, A. G., Eberhardt, M., Huber, J. B., & Sticht, H. (2013). An information-theoretic classification of amino acids for the assessment of interfaces in protein-protein docking. *Journal of Molecular Modeling*, *19*(9), 3901–3910. doi:10.1007/s00894-013-1916-7 PMID:23828247

Jayaram, B., Singh, T., Mukherjee, G., Mathur, A., Shekhar, S., & Shekhar, V. (2012). Sanjeevini: A freely accessible web-server for target directed lead molecule discovery. *BMC Bioinformatics*, *13*(Suppl 17), S7. doi:10.1186/1471-2105-13-S17-S7 PMID:23282245

Jerabek, P., Florian, J., Stiborova, M., & Martinek, V. (2014). Flexible docking-based molecular dynamics/steered molecular dynamics calculations of protein-protein contacts in a complex of cytochrome P450 1A2 with cytochrome b5. *Biochemistry*, *53*(42), 6695–6705. doi:10.1021/bi500814t PMID:25313797

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Jimenez-Garcia, B., Pons, C., & Fernandez-Recio, J. (2013). pyDockWEB: A web server for rigid-body protein-protein docking using electrostatics and desolvation scoring. *Bioinformatics (Oxford, England)*, *29*(13), 1698–1699. doi:10.1093/bioinformatics/btt262 PMID:23661696

Jimenez-Garcia, B., Pons, C., Svergun, D. I., Bernado, P., & Fernandez-Recio, J. (2015). pyDockSAXS: Protein-protein complex structure by SAXS and computational docking. *Nucleic Acids Research*, *43*(W1), W356-361. doi:10.1093/nar/gkv368 PMID:25897115

Jones, G., Willett, P., & Glen, R. C. (1995). Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *Journal of Molecular Biology*, *245*(1), 43–53. doi:10.1016/S0022-2836(95)80037-9 PMID:7823319

Jones, G., Willett, P., Glen, R. C., Leach, A. R., & Taylor, R. (1997). Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology*, *267*(3), 727–748. doi:10.1006/jmbi.1996.0897 PMID:9126849

Kantardjiev, A. A. (2011). GPU.proton.DOCK: Genuine Protein Ultrafast proton equilibria consistent DOCKing. *Nucleic Acids Res, 39*, W223-228. doi: 10.1093/nar/gkr412

Kantardjiev, A. A. (2012). Quantum.Ligand.Dock: protein-ligand docking with quantum entanglement refinement on a GPU system. *Nucleic Acids Res, 40*, W415-422. doi: 10.1093/nar/gks515

Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A. A., Aflalo, C., & Vakser, I. A. (1992). Molecular surface recognition: Determination of geometric fit between proteins and their ligands by correlation techniques. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(6), 2195–2199. doi:10.1073/pnas.89.6.2195 PMID:1549581

Khamis, M. A., Gomaa, W., & Ahmed, W. F. (2015). Machine learning in computational docking. *Artificial Intelligence in Medicine*, *63*(3), 135–152. doi:10.1016/j.artmed.2015.02.002 PMID:25724101

Kinnings, S. L., Liu, N., Tonge, P. J., Jackson, R. M., Xie, L., & Bourne, P. E. (2011). A machine learning-based method to improve docking scoring functions and its application to drug repurposing. *Journal of Chemical Information and Modeling*, *51*(2), 408–419. doi:10.1021/ci100369f PMID:21291174

Kitchen, D. B., Decornez, H., Furr, J. R., & Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nature Reviews. Drug Discovery*, *3*(11), 935–949. doi:10.1038/nrd1549 PMID:15520816

Konc, J., Konc, J. T., Penca, M., & Janezic, D. (2011). Binding-sites Prediction Assisting Protein-protein Docking. *Acta Chimica Slovenica*, *58*(3), 396–401. PMID:24062097

Kozakov, D., Brenke, R., Comeau, S. R., & Vajda, S. (2006). PIPER: An FFT-based protein docking program with pairwise potentials. *Proteins*, *65*(2), 392–406. doi:10.1002/prot.21117 PMID:16933295

Kramer, B., Rarey, M., & Lengauer, T. (1999). Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. *Proteins*, *37*(2), 228–241. doi:10.1002/(SICI)1097-0134(19991101)37:2<228::AID-PROT8>3.0.CO;2-8 PMID:10584068

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Krammer, A., Kirchhoff, P. D., Jiang, X., Venkatachalam, C. M., & Waldman, M. (2005). LigScore: A novel scoring function for predicting binding affinities. *Journal of Molecular Graphics & Modelling*, *23*(5), 395–407. doi:10.1016/j.jmgm.2004.11.007 PMID:15781182

Kroemer, R. T. (2003). Molecular modelling probes: Docking and scoring. *Biochemical Society Transactions*, *31*(Pt 5), 980–984. doi:10.1042/bst0310980 PMID:14505464

Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R., & Ferrin, T. E. (1982). A geometric approach to macromolecule-ligand interactions. *Journal of Molecular Biology*, *161*(2), 269–288. doi:10.1016/0022-2836(82)90153-X PMID:7154081

Kurcinski, M., Jamroz, M., Blaszczyk, M., Kolinski, A., & Kmiecik, S. (2015). CABS-dock web server for the flexible docking of peptides to proteins without prior knowledge of the binding site. *Nucleic Acids Research*, *43*(W1), W419-424. doi:10.1093/nar/gkv456 PMID:25943545

Labbe, C. M., Rey, J., Lagorce, D., Vavrusa, M., Becot, J., Sperandio, O., & Miteva, M. A. et al. (2015). MTiOpenScreen: A web server for structure-based virtual screening. *Nucleic Acids Research*, *43*(W1), W448-454. doi:10.1093/nar/gkv306 PMID:25855812

Lauck, F., Smith, C. A., Friedland, G. F., Humphris, E. L., & Kortemme, T. (2010). RosettaBackrub--a web server for flexible backbone protein structure modeling and design. *Nucleic Acids Res, 38*, W569-575. doi: 10.1093/nar/gkq369

Laurie, A. T., & Jackson, R. M. (2005). Q-SiteFinder: An energy-based method for the prediction of protein-ligand binding sites. *Bioinformatics (Oxford, England)*, *21*(9), 1908–1916. doi:10.1093/bioinformatics/bti315 PMID:15701681

Laurini, E., Col, V. D., Mamolo, M. G., Zampieri, D., Posocco, P., Fermeglia, M., & Pricl, S. et al. (2011). Homology Model and Docking-Based Virtual Screening for Ligands of the sigma1 Receptor. *ACS Med Chem Lett*, *2*(11), 834–839. doi:10.1021/ml2001505 PMID:24900272

Leach, A. R., Shoichet, B. K., & Peishoff, C. E. (2006). Prediction of protein-ligand interactions. Docking and scoring: Successes and gaps. *Journal of Medicinal Chemistry*, *49*(20), 5851–5855. doi:10.1021/jm060999m PMID:17004700

Lee, H., Heo, L., Lee, M. S., & Seok, C. (2015). GalaxyPepDock: A protein-peptide docking tool based on interaction similarity and energy optimization. *Nucleic Acids Research*, *43*(W1), W431-435. doi:10.1093/nar/gkv495 PMID:25969449

Lee, H. S., & Zhang, Y. (2012). BSP-SLIM: A blind low-resolution ligand-protein docking approach using predicted protein structures. *Proteins*, *80*(1), 93–110. doi:10.1002/prot.23165 PMID:21971880

Lee, P. H., Kuo, K. L., Chu, P. Y., Liu, E. M., & Lin, J. H. (2009). SLITHER: a web server for generating contiguous conformations of substrate molecules entering into deep active sites of proteins or migrating through channels in membrane transporters. *Nucleic Acids Res, 37*, W559-564. doi: 10.1093/nar/gkp359

Lesk, V. I., & Sternberg, M. J. (2008). 3D-Garden: A system for modelling protein-protein complexes based on conformational refinement of ensembles generated with the marching cubes algorithm. *Bioinformatics (Oxford, England)*, *24*(9), 1137–1144. doi:10.1093/bioinformatics/btn093 PMID:18326508

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Levinthal, C., Wodak, S. J., Kahn, P., & Dadivanian, A. K. (1975). Hemoglobin interaction in sickle cell fibers. I: Theoretical approaches to the molecular contacts. *Proceedings of the National Academy of Sciences of the United States of America*, *72*(4), 1330–1334. doi:10.1073/pnas.72.4.1330 PMID:1055409

Li, H., Gao, Z., Kang, L., Zhang, H., Yang, K., Yu, K.,... Jiang, H. (2006). TarFisDock: a web server for identifying drug targets with docking approach. *Nucleic Acids Res, 34*, W219-224. doi: 10.1093/nar/gkl114

Li, H., Leung, K. S., Ballester, P. J., & Wong, M. H. (2014). istar: A web platform for large-scale protein-ligand docking. *PLoS ONE*, *9*(1), e85678. doi:10.1371/journal.pone.0085678 PMID:24475049

Li, H., Zhang, H., Zheng, M., Luo, J., Kang, L., Liu, X., & Jiang, H. et al. (2009). An effective docking strategy for virtual screening based on multi-objective optimization algorithm. *BMC Bioinformatics*, *10*(1), 58. doi:10.1186/1471-2105-10-58 PMID:19210777

Li, L., Chen, R., & Weng, Z. (2003). RDOCK: Refinement of rigid-body protein docking predictions. *Proteins*, *53*(3), 693–707. doi:10.1002/prot.10460 PMID:14579360

Lill, M. A. (2011). Efficient incorporation of protein flexibility and dynamics into molecular docking simulations. *Biochemistry*, *50*(28), 6157–6169. doi:10.1021/bi2004558 PMID:21678954

Lindahl, E., & Delarue, M. (2005). Refinement of docked protein-ligand and protein-DNA structures using low frequency normal mode amplitude optimization. *Nucleic Acids Research*, *33*(14), 4496–4506. doi:10.1093/nar/gki730 PMID:16087736

Liu, M., & Wang, S. (1999). MCDOCK: A Monte Carlo simulation approach to the molecular docking problem. *Journal of Computer-Aided Molecular Design*, *13*(5), 435–451. doi:10.1023/A:1008005918983 PMID:10483527

Liu, X., Ouyang, S., Yu, B., Liu, Y., Huang, K., Gong, J.,... Jiang, H. (2010). PharmMapper server: a web server for potential drug target identification using pharmacophore mapping approach. *Nucleic Acids Res, 38*, W609-614. doi: 10.1093/nar/gkq300

Liu, Y., Zhao, L., Li, W., Zhao, D., Song, M., & Yang, Y. (2013). FIPSDock: A new molecular docking technique driven by fully informed swarm optimization algorithm. *Journal of Computational Chemistry*, *34*(1), 67–75. doi:10.1002/jcc.23108 PMID:22961860

London, N., Miller, R. M., Krishnan, S., Uchida, K., Irwin, J. J., Eidam, O., & Taunton, J. et al. (2014). Covalent docking of large libraries for the discovery of chemical probes. *Nature Chemical Biology*, *10*(12), 1066–1072. doi:10.1038/nchembio.1666 PMID:25344815

Lyskov, S., Chou, F. C., Conchuir, S. O., Der, B. S., Drew, K., Kuroda, D., & Das, R. et al. (2013). Serverification of molecular modeling applications: The Rosetta Online Server that Includes Everyone (ROSIE). *PLoS ONE*, *8*(5), e63906. PMID:23717507

Lyskov, S., & Gray, J. J. (2008). The RosettaDock server for local protein-protein docking. *Nucleic Acids Res, 36*, W233-238. doi: 10.1093/nar/gkn216

Macindoe, G., Mavridis, L., Venkatraman, V., Devignes, M. D., & Ritchie, D. W. (2010). HexServer: an FFT-based protein docking server powered by graphics processors. *Nucleic Acids Res, 38*, W445-449. doi: 10.1093/nar/gkq311

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Madeira, F., Tinti, M., Murugesan, G., Berrett, E., Stafford, M., Toth, R., & Barton, G. J. et al. (2015). 14-3-3-Pred: Improved methods to predict 14-3-3-binding phosphopeptides. *Bioinformatics (Oxford, England)*, *31*(14), 2276–2283. doi:10.1093/bioinformatics/btv133 PMID:25735772

Malhotra, S., Mathew, O. K., & Sowdhamini, R. (2015). DOCKSCORE: A webserver for ranking protein-protein docked poses. *BMC Bioinformatics*, *16*(1), 127. doi:10.1186/s12859-015-0572-6 PMID:25902779

Mandal, S., Moudgil, M., & Mandal, S. K. (2009). Rational drug design. *European Journal of Pharmacology*, *625*(1-3), 90–100. doi:10.1016/j.ejphar.2009.06.065 PMID:19835861

Martin, L., Catherinot, V., & Labesse, G. (2006). kinDOCK: a tool for comparative docking of protein kinase ligands. *Nucleic Acids Res, 34*, W325-329. doi: 10.1093/nar/gkl211

Matsuno, K., Masuda, Y., Uehara, Y., Sato, H., Muroya, A., Takahashi, O., & Asai, A. et al. (2010). Identification of a new series of STAT3 inhibitors by virtual screening. *ACS Med Chem Lett*, *1*(8), 371–375. doi:10.1021/ml1000273 PMID:24900220

McMartin, C., & Bohacek, R. S. (1997). QXP: Powerful, rapid computer algorithms for structure-based drug design. *Journal of Computer-Aided Molecular Design*, *11*(4), 333–344. doi:10.1023/A:1007907728892 PMID:9334900

McQuarrie, D. A. (1976). *StatisticalMechanics*. New York: Harper Collins Publishers.

Meier, R., Pippel, M., Brandt, F., Sippl, W., & Baldauf, C. (2010). ParaDockS: A framework for molecular docking with population-based metaheuristics. *Journal of Chemical Information and Modeling*, *50*(5), 879–889. doi:10.1021/ci900467x PMID:20415499

Mendez, R., Leplae, R., De Maria, L., & Wodak, S. J. (2003). Assessment of blind predictions of protein-protein interactions: Current status of docking methods. *Proteins*, *52*(1), 51–67. doi:10.1002/prot.10393 PMID:12784368

Mendez, R., Leplae, R., Lensink, M. F., & Wodak, S. J. (2005). Assessment of CAPRI predictions in rounds 3-5 shows progress in docking procedures. *Proteins*, *60*(2), 150–169. doi:10.1002/prot.20551 PMID:15981261

Miller, M. D., Kearsley, S. K., Underwood, D. J., & Sheridan, R. P. (1994). FLOG: A system to select 'quasi-flexible' ligands complementary to a receptor of known three-dimensional structure. *Journal of Computer-Aided Molecular Design*, *8*(2), 153–174. doi:10.1007/BF00119865 PMID:8064332

Miteva, M. A., Lee, W. H., Montes, M. O., & Villoutreix, B. O. (2005). Fast structure-based virtual ligand screening combining FRED, DOCK, and Surflex. *Journal of Medicinal Chemistry*, *48*(19), 6012–6022. doi:10.1021/jm050262h PMID:16162004

Mitra, P., & Pal, D. (2011). PRUNE and PROBE--two modular web services for protein-protein docking. *Nucleic Acids Res, 39*, W229-234. doi: 10.1093/nar/gkr317

Mizutani, M. Y., Tomioka, N., & Itai, A. (1994). Rational automatic search method for stable docking models of protein and ligand. *Journal of Molecular Biology*, *243*(2), 310–326. doi:10.1006/jmbi.1994.1656 PMID:7932757

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Moal, I. H., Dapkunas, J., & Fernandez-Recio, J. (2015). Inferring the microscopic surface energy of protein-protein interfaces from mutation data. *Proteins*, *83*(4), 640–650. doi:10.1002/prot.24761 PMID:25586563

Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., & Olson, A. J. (1999). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry*, *19*(14), 1639–1662. doi:10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B

Morris, G. M., Goodsell, D. S., Huey, R., & Olson, A. J. (1996). Distributed automated docking of flexible ligands to proteins: Parallel applications of AutoDock 2.4. *Journal of Computer-Aided Molecular Design*, *10*(4), 293–304. doi:10.1007/BF00124499 PMID:8877701

Morris, G. M., Green, L. G., Radic, Z., Taylor, P., Sharpless, K. B., Olson, A. J., & Grynszpan, F. (2013). Automated docking with protein flexibility in the design of femtomolar "click chemistry" inhibitors of acetylcholinesterase. *Journal of Chemical Information and Modeling*, *53*(4), 898–906. doi:10.1021/ci300545a PMID:23451944

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, *30*(16), 2785–2791. doi:10.1002/jcc.21256 PMID:19399780

Morris, G. M., & Lim-Wilby, M. (2008). Molecular docking. In A. Kukol (Ed.), *Molecular Modeling of Proteins* (Vol. 443, pp. 365–382). Totowa: Humana Press. doi:10.1007/978-1-59745-177-2_19

Mottarella, S. E., Beglov, D., Beglova, N., Nugent, M. A., Kozakov, D., & Vajda, S. (2014). Docking server for the identification of heparin binding sites on proteins. *Journal of Chemical Information and Modeling*, *54*(7), 2068–2078. doi:10.1021/ci500115j PMID:24974889

Muegge, I., & Enyedy, I. (2003). Virtual screening. In D. J. Abraham (Ed.), *Burger's Medicinal Chemistry and Drug Discovery, Drug Discovery* (Vol. 1, pp. 243–279). New Jersey: John Wiley & Sons. doi:10.1002/0471266949.bmc006

Muegge, I., & Martin, Y. C. (1999). A general and fast scoring function for protein-ligand interactions: A simplified potential approach. *Journal of Medicinal Chemistry*, *42*(5), 791–804. doi:10.1021/jm980536j PMID:10072678

Mukherjee, G., Patra, N., Barua, P., & Jayaram, B. (2011). A fast empirical GAFF compatible partial atomic charge assignment scheme for modeling interactions of small molecules with biomolecular targets. *Journal of Computational Chemistry*, *32*(5), 893–907. doi:10.1002/jcc.21671 PMID:21341292

Murray, C. W., Baxter, C. A., & Frenkel, A. D. (1999). The sensitivity of the results of molecular docking to induced fit effects: Application to thrombin, thermolysin and neuraminidase. *Journal of Computer-Aided Molecular Design*, *13*(6), 547–562. doi:10.1023/A:1008015827877 PMID:10584214

Nabuurs, S. B., Wagener, M., & de Vlieg, J. (2007). A flexible approach to induced fit docking. *Journal of Medicinal Chemistry*, *50*(26), 6507–6518. doi:10.1021/jm070593p PMID:18031000

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Namasivayam, V., & Gunther, R. (2007). pso@autodock: A fast flexible molecular docking program based on Swarm intelligence. *Chemical Biology & Drug Design*, *70*(6), 475–484. doi:10.1111/j.1747-0285.2007.00588.x PMID:17986206

Ng, M. C., Fong, S., & Siu, S. W. (2015). PSOVina: The hybrid particle swarm optimization algorithm for protein-ligand docking. *Journal of Bioinformatics and Computational Biology*, *13*(3), 1541007. doi:10.1142/S0219720015410073 PMID:25800162

Nicolaou, C. A., & Brown, N. (2013). Multi-objective optimization methods in drug design. *Drug Discovery Today. Technologies*, *10*(3), e427–e435. doi:10.1016/j.ddtec.2013.02.001 PMID:24050140

Nisius, B., & Gohlke, H. (2012). Alignment-independent comparison of binding sites based on DrugScore potential fields encoded by 3D Zernike descriptors. *Journal of Chemical Information and Modeling*, *52*(9), 2339–2347. doi:10.1021/ci300244y PMID:22881363

Nolan, T. L., Lapinsky, D. J., Talbot, J. N., Indarte, M., Liu, Y., Manepalli, S., & Surratt, C. K. et al. (2011). Identification of a novel selective serotonin reuptake inhibitor by coupling monoamine transporter-based virtual screening and rational molecular hybridization. *ACS Chemical Neuroscience*, *2*(9), 544–552. doi:10.1021/cn200044x PMID:21966587

Olimpieri, P. P., Chailyan, A., Tramontano, A., & Marcatili, P. (2013). Prediction of site-specific interactions in antibody-antigen complexes: The proABC method and server. *Bioinformatics (Oxford, England)*, *29*(18), 2285–2291. doi:10.1093/bioinformatics/btt369 PMID:23803466

Omori, S., & Kitao, A. (2013). CyClus: A fast, comprehensive cylindrical interface approximation clustering/reranking method for rigid-body protein-protein docking decoys. *Proteins*, *81*(6), 1005–1016. doi:10.1002/prot.24252 PMID:23344972

Ou-Yang, S. S., Lu, J. Y., Kong, X. Q., Liang, Z. J., Luo, C., & Jiang, H. (2012). Computational drug discovery. *Acta Pharmacologica Sinica*, *33*(9), 1131–1140. doi:10.1038/aps.2012.109 PMID:22922346

Ouyang, X., Zhou, S., Ge, Z., Li, R., & Kwoh, C. K. (2013). CovalentDock Cloud: a web server for automated covalent docking. *Nucleic Acids Res, 41*, W329-332.

Pak, Y., & Wang, S. (2000). Application of a Molecular Dynamics Simulation Method with a Generalized Effective Potential to the Flexible Molecular Docking Problems. *The Journal of Physical Chemistry B*, *104*(2), 354–359. doi:10.1021/jp993073h

Pala, N., Stevaert, A., Dallocchio, R., Dessi, A., Rogolino, D., Carcelli, M.,... Naesens, L. (2015). Virtual screening and biological validation of novel influenza virus PA endonuclease inhibitors. *ACS Medicinal Chemistry Letters ACS Med. Chem. Lett.*

Patel, P. D., Patel, M. R., Kaushik-Basu, N., & Talele, T. T. (2008). 3D QSAR and molecular docking studies of benzimidazole derivatives as hepatitis C virus NS5B polymerase inhibitors. *Journal of Chemical Information and Modeling*, *48*(1), 42–55. doi:10.1021/ci700266z PMID:18076152

Pevzner, Y., Frugier, E., Schalk, V., Caflisch, A., & Woodcock, H. L. (2014). Fragment-based docking: Development of the CHARMMing Web user interface as a platform for computer-aided drug design. *Journal of Chemical Information and Modeling*, *54*(9), 2612–2620. doi:10.1021/ci500322k PMID:25151852

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Philips, A., Milanowska, K., Lach, G., & Bujnicki, J. M. (2013). LigandRNA: Computational predictor of RNA-ligand interactions. *RNA (New York, N.Y.)*, *19*(12), 1605–1616. doi:10.1261/rna.039834.113 PMID:24145824

Pierce, B., Tong, W., & Weng, Z. (2005). M-ZDOCK: A grid-based approach for Cn symmetric multimer docking. *Bioinformatics (Oxford, England)*, *21*(8), 1472–1478. doi:10.1093/bioinformatics/bti229 PMID:15613396

Pierce, B. G., Hourai, Y., & Weng, Z. (2011). Accelerating protein docking in ZDOCK using an advanced 3D convolution library. *PLoS ONE*, *6*(9), e24657. doi:10.1371/journal.pone.0024657 PMID:21949741

Pierce, B. G., Wiehe, K., Hwang, H., Kim, B. H., Vreven, T., & Weng, Z. (2014). ZDOCK server: Interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics (Oxford, England)*, *30*(12), 1771–1773. doi:10.1093/bioinformatics/btu097 PMID:24532726

Pirhadi, S., Shiri, F., & Ghasemi, J. B. (2013). Methods and applications of structure based pharmacophores in drug discovery. *Current Topics in Medicinal Chemistry*, *13*(9), 1036–1047. doi:10.2174/1568026611313090006 PMID:23651482

Plewczynski, D., Lazniewski, M., von Grotthuss, M., Rychlewski, L., & Ginalski, K. (2011). VoteDock: Consensus docking method for prediction of protein-ligand interactions. *Journal of Computational Chemistry*, *32*(4), 568–581. doi:10.1002/jcc.21642 PMID:20812324

Politi, A., Durdagi, S., Moutevelis-Minakakis, P., Kokotos, G., & Mavromoustakos, T. (2010). Development of accurate binding affinity predictions of novel renin inhibitors through molecular docking studies. *Journal of Molecular Graphics & Modelling*, *29*(3), 425–435. doi:10.1016/j.jmgm.2010.08.003 PMID:20855222

Pons, C., Jimenez-Gonzalez, D., Gonzalez-Alvarez, C., Servat, H., Cabrera-Benitez, D., Aguilar, X., & Fernandez-Recio, J. (2012). Cell-Dock: High-performance protein-protein docking. *Bioinformatics (Oxford, England)*, *28*(18), 2394–2396. doi:10.1093/bioinformatics/bts454 PMID:22815362

Pons, C., Solernou, A., Perez-Cano, L., Grosdidier, S., & Fernandez-Recio, J. (2010). Optimization of pyDock for the new CAPRI challenges: Docking of homology-based models, domain-domain assembly and protein-RNA binding. *Proteins*, *78*(15), 3182–3188. doi:10.1002/prot.22773 PMID:20602351

Pons, J. L., & Labesse, G. (2009). @TOME-2: a new pipeline for comparative modeling of protein-ligand complexes. *Nucleic Acids Res, 37*, W485-491. doi: 10.1093/nar/gkp368

Porollo, A., & Meller, J. (2010). POLYVIEW-MM: web-based platform for animation and analysis of molecular simulations. *Nucleic Acids Res, 38*, W662-666. doi: 10.1093/nar/gkq445

Purisima, E. O., & Hogues, H. (2012). Protein-ligand binding free energies from exhaustive docking. *The Journal of Physical Chemistry B*, *116*(23), 6872–6879. doi:10.1021/jp212646s PMID:22432509

Raub, S., Steffen, A., Kamper, A., & Marian, C. M. (2008). AIScore chemically diverse empirical scoring function employing quantum chemical binding energies of hydrogen-bonded complexes. *Journal of Chemical Information and Modeling*, *48*(7), 1492–1510. doi:10.1021/ci7004669 PMID:18597446

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Ren, J., Williams, N., Clementi, L., Krishnan, S., & Li, W. W. (2010). Opal web services for biomedical applications. *Nucleic Acids Res, 38*, W724-731. doi: 10.1093/nar/gkq503

Ricci, C. G., & Netz, P. A. (2009). Docking studies on DNA-ligand interactions: Building and application of a protocol to identify the binding mode. *Journal of Chemical Information and Modeling*, *49*(8), 1925–1935. doi:10.1021/ci9001537 PMID:19655805

Ritchie, D. W., & Kemp, G. J. (2000). Protein docking using spherical polar Fourier correlations. *Proteins*, *39*(2), 178–194. doi:10.1002/(SICI)1097-0134(20000501)39:2<178::AID-PROT8>3.0.CO;2-6 PMID:10737939

Ritchie, D. W., Kozakov, D., & Vajda, S. (2008). Accelerating and focusing protein-protein docking correlations using multi-dimensional rotational FFT generating functions. *Bioinformatics (Oxford, England)*, *24*(17), 1865–1873. doi:10.1093/bioinformatics/btn334 PMID:18591193

Ritchie, D. W., & Venkatraman, V. (2010). Ultra-fast FFT protein docking on graphics processors. *Bioinformatics (Oxford, England)*, *26*(19), 2398–2405. doi:10.1093/bioinformatics/btq444 PMID:20685958

Ritchie, D. W., Venkatraman, V., & Mavridis, L. (2010). Using graphics processors to accelerate protein docking calculations. *Studies in Health Technology and Informatics*, *159*, 146–155. PMID:20543434

Rueda, M., Bottegoni, G., & Abagyan, R. (2009). Consistent improvement of cross-docking results using binding site ensembles generated with elastic network normal modes. *Journal of Chemical Information and Modeling*, *49*(3), 716–725. doi:10.1021/ci8003732 PMID:19434904

Saladin, A., Rey, J., Thevenet, P., Zacharias, M., Moroy, G., & Tuffery, P. (2014). PEP-SiteFinder: a tool for the blind identification of peptide binding sites on protein surfaces. *Nucleic Acids Res, 42*, W221-226.

Salemme, F. R. (1976). An hypothetical structure for an intermolecular electron transfer complex of cytochromes c and b5. *Journal of Molecular Biology*, *102*(3), 563–568. doi:10.1016/0022-2836(76)90334-X PMID:178879

Sandal, M., Duy, T. P., Cona, M., Zung, H., Carloni, P., Musiani, F., & Giorgetti, A. (2013). GOMoDo: A GPCRs online modeling and docking webserver. *PLoS ONE*, *8*(9), e74092. doi:10.1371/journal.pone.0074092 PMID:24058518

Schellhammer, I., & Rarey, M. (2004). FlexX-Scan: Fast, structure-based virtual screening. *Proteins*, *57*(3), 504–517. doi:10.1002/prot.20217 PMID:15382244

Schmidt, T., Bergner, A., & Schwede, T. (2014). Modelling three-dimensional protein structures for applications in drug design. *Drug Discovery Today*, *19*(7), 890–897. doi:10.1016/j.drudis.2013.10.027 PMID:24216321

Schneider, G., & Fechner, U. (2005). Computer-based de novo design of drug-like molecules. *Nature Reviews. Drug Discovery*, *4*(8), 649–663. doi:10.1038/nrd1799 PMID:16056391

Shimoda, T., Suzuki, S., Ohue, M., Ishida, T., & Akiyama, Y. (2015). Protein-protein docking on hardware accelerators: Comparison of GPU and MIC architectures. *BMC Systems Biology*, *9*(Suppl 1), S6. doi:10.1186/1752-0509-9-S1-S6 PMID:25707855

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Shirvanyants, D., Alexandrova, A. N., & Dokholyan, N. V. (2011). Rigid substructure search. *Bioinformatics (Oxford, England)*, *27*(9), 1327–1329. doi:10.1093/bioinformatics/btr129 PMID:21460026

Shoemaker, B. A., Zhang, D., Tyagi, M., Thangudu, R. R., Fong, J. H., Marchler-Bauer, A., & Panchenko, A. R. et al. (2012). IBIS (Inferred Biomolecular Interaction Server) reports, predicts and integrates multiple types of conserved interactions for proteins. *Nucleic Acids Research*, *40*(Database issue), D834–D840. doi:10.1093/nar/gkr997 PMID:22102591

Singh, T., Biswas, D., & Jayaram, B. (2011). AADS--an automated active site identification, docking, and scoring protocol for protein targets based on physicochemical descriptors. *Journal of Chemical Information and Modeling*, *51*(10), 2515–2527. doi:10.1021/ci200193z PMID:21877713

Sircar, A., Kim, E. T., & Gray, J. J. (2009). RosettaAntibody: antibody variable region homology modeling server. *Nucleic Acids Res, 37*, W474-479. doi: 10.1093/nar/gkp387

Sondergaard, C. R., Garrett, A. E., Carstensen, T., Pollastri, G., & Nielsen, J. E. (2009). Structural artifacts in protein-ligand X-ray structures: Implications for the development of docking scoring functions. *Journal of Medicinal Chemistry*, *52*(18), 5673–5684. doi:10.1021/jm8016464 PMID:19711919

Sotriffer, C., Klebe, G., Stahl, M., & Bohm, H.-J. (2003). Docking and scoring functions/Virtual screening. In D. J. Abraham (Ed.), *Burger's Medicinal Chemistry and Drug Discovery, Drug Discovery* (Vol. 1, pp. 281–416). John Wiley & Sons. doi:10.1002/0471266949.bmc007

Sotriffer, C. A., Gohlke, H., & Klebe, G. (2002). Docking into knowledge-based potential fields: A comparative evaluation of DrugScore. *Journal of Medicinal Chemistry*, *45*(10), 1967–1970. doi:10.1021/jm025507u PMID:11985464

Sousa, S. F., Fernandes, P. A., & Ramos, M. J. (2006). Protein-ligand docking: Current status and future challenges. *Proteins*, *65*(1), 15–26. doi:10.1002/prot.21082 PMID:16862531

Sousa, S. F., Ribeiro, A. J., Coimbra, J. T., Neves, R. P., Martins, S. A., Moorthy, N. S., & Ramos, M. J. et al. (2013). Protein-ligand docking in the new millennium--a retrospective of 10 years in the field. *Current Medicinal Chemistry*, *20*(18), 2296–2314. doi:10.2174/09298673113201800002 PMID:23531220

Spitzer, R., & Jain, A. N. (2012). Surflex-Dock: Docking benchmarks and real-world application. *Journal of Computer-Aided Molecular Design*, *26*(6), 687–699. doi:10.1007/s10822-011-9533-y PMID:22569590

Sukhwani, B., & Herbordt, M. C. (2010). FPGA acceleration of rigid-molecule docking codes. *IET Comput Digit Tech*, *4*(3), 184–195. doi:10.1049/iet-cdt.2009.0013 PMID:21857870

Tan, K. P., Varadarajan, R., & Madhusudhan, M. S. (2011). DEPTH: a web server to compute depth and predict small-molecule binding cavities in proteins. *Nucleic Acids Res, 39*, W242-248. doi: 10.1093/nar/gkr356

Taylor, J. S., & Burnett, R. M. (2000). DARWIN: A program for docking flexible molecules. *Proteins*, *41*(2), 173–191. doi:10.1002/1097-0134(20001101)41:2<173::AID-PROT30>3.0.CO;2-3 PMID:10966571

Taylor, R. D., Jewsbury, P. J., & Essex, J. W. (2002). A review of protein-small molecule docking methods. *Journal of Computer-Aided Molecular Design*, *16*(3), 151–166. doi:10.1023/A:1020155510718 PMID:12363215

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Terashi, G., Takeda-Shitaka, M., Kanou, K., Iwadate, M., Takaya, D., & Umeyama, H. (2007). The SKE-DOCK server and human teams based on a combined method of shape complementarity and free energy estimation. *Proteins*, *69*(4), 866–872. doi:10.1002/prot.21772 PMID:17853449

Tietze, S., & Apostolakis, J. (2007). GlamDock: Development and validation of a new docking tool on several thousand protein-ligand complexes. *Journal of Chemical Information and Modeling*, *47*(4), 1657–1672. doi:10.1021/ci7001236 PMID:17585857

Tintori, C., Magnani, M., Schenone, S., & Botta, M. (2009). Docking, 3D-QSAR studies and in silico ADME prediction on c-Src tyrosine kinase inhibitors. *European Journal of Medicinal Chemistry*, *44*(3), 990–1000. doi:10.1016/j.ejmech.2008.07.002 PMID:18722033

Tondel, K., Anderssen, E., & Drablos, F. (2006). Protein Alpha Shape (PAS) Dock: A new gaussian-based score function suitable for docking in homology modelled protein structures. *Journal of Computer-Aided Molecular Design*, *20*(3), 131–144. doi:10.1007/s10822-006-9041-7 PMID:16652207

Torchala, M., & Bates, P. A. (2014). Predicting the structure of protein-protein complexes using the SwarmDock Web Server. *Methods in Molecular Biology (Clifton, N.J.)*, *1137*, 181–197. doi:10.1007/978-1-4939-0366-5_13 PMID:24573482

Torchala, M., Moal, I. H., Chaleil, R. A., Fernandez-Recio, J., & Bates, P. A. (2013). SwarmDock: A server for flexible protein-protein docking. *Bioinformatics (Oxford, England)*, *29*(6), 807–809. doi:10.1093/bioinformatics/btt038 PMID:23343604

Tovchigrechko, A., & Vakser, I. A. (2006). GRAMM-X public web server for protein-protein docking. *Nucleic Acids Res, 34*, W310-314. doi: 10.1093/nar/gkl206

Trosset, J. Y., & Scheraga, H. A. (1998). Reaching the global minimum in docking simulations: A Monte Carlo energy minimization approach using Bezier splines. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(14), 8011–8015. doi:10.1073/pnas.95.14.8011 PMID:9653131

Tsai, T. Y., Chang, K. W., & Chen, C. Y. (2011). iScreen: World's first cloud-computing web server for virtual screening and de novo drug design based on TCM database@Taiwan. *Journal of Computer-Aided Molecular Design*, *25*(6), 525–531. doi:10.1007/s10822-011-9438-9 PMID:21647737

Tuccinardi, T., Nuti, E., Ortore, G., Supuran, C. T., Rossello, A., & Martinelli, A. (2007). Analysis of human carbonic anhydrase II: Docking reliability and receptor-based 3D-QSAR study. *Journal of Chemical Information and Modeling*, *47*(2), 515–525. doi:10.1021/ci600469w PMID:17295464

Tuffery, P., & Derreumaux, P. (2012). Flexibility and binding affinity in protein-ligand, protein-protein and multi-component protein interactions: Limitations of current computational approaches. *Journal of the Royal Society, Interface*, *9*(66), 20–33. PMID:21993006

Tuszynska, I., Magnus, M., Jonak, K., Dawson, W., & Bujnicki, J. M. (2015). NPDock: A web server for protein-nucleic acid docking. *Nucleic Acids Research*, *43*(W1), W425-430. doi:10.1093/nar/gkv493 PMID:25977296

Vajda, S., Hall, D. R., & Kozakov, D. (2013). Sampling and scoring: A marriage made in heaven. *Proteins*, *81*(11), 1874–1884. doi:10.1002/prot.24343 PMID:23775627

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

van Dijk, A. D., & Bonvin, A. M. (2006). Solvated docking: Introducing water into the modelling of biomolecular complexes. *Bioinformatics (Oxford, England)*, *22*(19), 2340–2347. doi:10.1093/bioinformatics/btl395 PMID:16899489

Velec, H. F., Gohlke, H., & Klebe, G. (2005). DrugScore(CSD)-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. *Journal of Medicinal Chemistry*, *48*(20), 6296–6303. doi:10.1021/jm050436v PMID:16190756

Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W., & Taylor, R. D. (2003). Improved protein-ligand docking using GOLD. *Proteins*, *52*(4), 609–623. doi:10.1002/prot.10465 PMID:12910460

Veselovsky, A. V., & Ivanov, A. S. (2003). Strategy of computer-aided drug design. *Current Drug Targets. Infectious Disorders*, *3*(1), 33–40. doi:10.2174/1568005033342145 PMID:12570731

Viswanath, S., Ravikant, D. V., & Elber, R. (2014). DOCK/PIERR: Web server for structure prediction of protein-protein complexes. *Methods in Molecular Biology (Clifton, N.J.)*, *1137*, 199–207. doi:10.1007/978-1-4939-0366-5_14 PMID:24573483

Vitorovic-Todorovic, M. D., Cvijetic, I. N., Juranic, I. O., & Drakulic, B. J. (2012). The 3D-QSAR study of 110 diverse, dual binding, acetylcholinesterase inhibitors based on alignment independent descriptors (GRIND-2). The effects of conformation on predictive power and interpretability of the models. *Journal of Molecular Graphics & Modelling*, *38*, 194–210. doi:10.1016/j.jmgm.2012.08.001 PMID:23073222

Vreven, T., Pierce, B. G., Hwang, H., & Weng, Z. (2009). Performance of ZDOCK in CAPRI rounds 20-26. *Proteomics*, *9*(12), 2175–2182. doi:10.1002/pmic.200800646 PMID:24123140

Vreven, T., Pierce, B. G., Hwang, H., & Weng, Z. (2013). Performance of ZDOCK in CAPRI rounds 20-26. *Proteins*, *81*(12), 2175–2182. doi:10.1002/prot.24432 PMID:24123140

Wang, J. C., Chu, P. Y., Chen, C. M., & Lin, J. H. (2012). idTarget: a web server for identifying protein targets of small chemical molecules with robust scoring functions and a divide-and-conquer docking approach. *Nucleic Acids Res, 40*, W393-399.

Wei, Y., Li, J., Chen, Z., Wang, F., Huang, W., Hong, Z., & Lin, J. (2015). Multistage virtual screening and identification of novel HIV-1 protease inhibitors by integrating SVM, shape, pharmacophore and docking methods. *European Journal of Medicinal Chemistry*, *101*, 409–418. doi:10.1016/j.ejmech.2015.06.054 PMID:26185005

Weiner, S. J., Kollman, P. A., Nguyen, D. T., & Case, D. A. (1986). An all atom force field for simulations of proteins and nucleic acids. *Journal of Computational Chemistry*, *7*(2), 230–252. doi:10.1002/jcc.540070216

Welch, W., Ruppert, J., & Jain, A. N. (1996). Hammerhead: Fast, fully automated docking of flexible ligands to protein binding sites. *Chemistry & Biology*, *3*(6), 449–462. doi:10.1016/S1074-5521(96)90093-9 PMID:8807875

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Wiehe, K., Pierce, B., Mintseris, J., Tong, W. W., Anderson, R., Chen, R., & Weng, Z. (2005). ZDOCK and RDOCK performance in CAPRI rounds 3, 4, and 5. *Proteins*, *60*(2), 207–213. doi:10.1002/prot.20559 PMID:15981263

Wiehe, K., Pierce, B., Tong, W. W., Hwang, H., Mintseris, J., & Weng, Z. (2007). The performance of ZDOCK and ZRANK in rounds 6-11 of CAPRI. *Proteins*, *69*(4), 719–725. doi:10.1002/prot.21747 PMID:17803212

Wilson, G. L., & Lill, M. A. (2011). Integrating structure-based and ligand-based approaches for computational drug design. *Future Med Chem*, *3*(6), 735–750. doi:10.4155/fmc.11.18 PMID:21554079

Wu, J., Hong, B., Takeda, T., & Guo, J. T. (2012). High performance transcription factor-DNA docking with GPU computing. *Proteome Science*, *10*(Suppl 1), S17. doi:10.1186/1477-5956-10-S1-S17 PMID:22759575

Xu, B., Yang, Y., Liang, H., & Zhou, Y. (2009). An all-atom knowledge-based energy function for protein-DNA threading, docking decoy discrimination, and prediction of transcription-factor binding profiles. *Proteins*, *76*(3), 718–730. doi:10.1002/prot.22384 PMID:19274740

Xu, D., Jaroszewski, L., Li, Z., & Godzik, A. (2014). AIDA: ab initio domain assembly server. *Nucleic Acids Res, 42*, W308-313.

Xue, L. C., Jordan, R. A., El-Manzalawy, Y., Dobbs, D., & Honavar, V. (2014). DockRank: Ranking docked conformations using partner-specific sequence homology-based protein interface prediction. *Proteins*, *82*(2), 250–267. doi:10.1002/prot.24370 PMID:23873600

Yakar, R., & Akten, E. D. (2014). Discovery of high affinity ligands for beta2-adrenergic receptor through pharmacophore-based high-throughput virtual screening and docking. *Journal of Molecular Graphics & Modelling*, *53*, 148–160. doi:10.1016/j.jmgm.2014.07.007 PMID:25137647

Yang, C. Y., Sun, H., Chen, J., Nikolovska-Coleska, Z., & Wang, S. (2009). Importance of ligand reorganization free energy in protein-ligand binding-affinity prediction. *Journal of the American Chemical Society*, *131*(38), 13709–13721. doi:10.1021/ja9039373 PMID:19736924

Yang, Y., Zhao, H., Wang, J., & Zhou, Y. (2014). SPOT-Seq-RNA: Predicting protein-RNA complex structure and RNA-binding function by fold recognition and binding affinity prediction. *Methods in Molecular Biology (Clifton, N.J.)*, *1137*, 119–130. doi:10.1007/978-1-4939-0366-5_9 PMID:24573478

Yi, H., Qiu, S., Cao, Z., Wu, Y., & Li, W. (2008). Molecular basis of inhibitory peptide maurotoxin recognizing Kv1.2 channel explored by ZDOCK and molecular dynamic simulations. *Proteins*, *70*(3), 844–854. doi:10.1002/prot.21706 PMID:17729277

Yin, S., Biedermannova, L., Vondrasek, J., & Dokholyan, N. V. (2008). MedusaScore: An accurate force field-based scoring function for virtual drug screening. *Journal of Chemical Information and Modeling*, *48*(8), 1656–1662. doi:10.1021/ci8001167 PMID:18672869

Yu, J., Zhou, Y., Tanaka, I., & Yao, M. (2010). Roll: A new algorithm for the detection of protein pockets and cavities with a rolling probe sphere. *Bioinformatics (Oxford, England)*, *26*(1), 46–52. doi:10.1093/bioinformatics/btp599 PMID:19846440

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Yu, X., Martinez, M., Gable, A. L., Fuller, J. C., Bruce, N. J., Richter, S., & Wade, R. C. (2015). web-SDA: A web server to simulate macromolecular diffusional association. *Nucleic Acids Research*, *43*(W1), W220-224. doi:10.1093/nar/gkv335 PMID:25883142

Yuriev, E., Holien, J., & Ramsland, P. A. (2015). Improvements, trends, and new ideas in molecular docking: 2012-2013 in review. *Journal of Molecular Recognition*, *28*(10), 581–604. doi:10.1002/jmr.2471 PMID:25808539

Zhang, C., & Lai, L. (2011). SDOCK: A global protein-protein docking program using stepwise force-field potentials. *Journal of Computational Chemistry*, *32*(12), 2598–2612. doi:10.1002/jcc.21839 PMID:21618559

Zhang, C., Liu, S., & Zhou, Y. (2005). Docking prediction using biological information, ZDOCK sampling technique, and clustering guided by the DFIRE statistical energy function. *Proteins*, *60*(2), 314–318. doi:10.1002/prot.20576 PMID:15981255

Zhou, H., & Skolnick, J. (2013). FINDSITE(comb): A threading/structure-based, proteomic-scale virtual ligand screening approach. *Journal of Chemical Information and Modeling*, *53*(1), 230–240. doi:10.1021/ci300510n PMID:23240691

Zhou, Z., & Madura, J. D. (2004). CoMFA 3D-QSAR analysis of HIV-1 RT nonnucleoside inhibitors, TIBO derivatives based on docking conformation and alignment. *Journal of Chemical Information and Computer Sciences*, *44*(6), 2167–2178. doi:10.1021/ci049893v PMID:15554687

Zsila, F., Bikadi, Z., Malik, D., Hari, P., Pechan, I., Berces, A., & Hazai, E. (2011). Evaluation of drug-human serum albumin binding interactions with support vector machine aided online automated docking. *Bioinformatics (Oxford, England)*, *27*(13), 1806–1813. doi:10.1093/bioinformatics/btr284 PMID:21593135

Zsoldos, Z., Reid, D., Simon, A., Sadjad, B. S., & Johnson, A. P. (2006). eHiTS: An innovative approach to the docking and scoring function problems. *Current Protein & Peptide Science*, *7*(5), 421–435. doi:10.2174/138920306778559412 PMID:17073694

KEY TERMS AND DEFINITIONS

Induced-Fit Concept: Conformational changes of both the receptor and ligand induced during the binding process.

Ligand: Molecule able to bind to another molecule (usually a macromolecule).

Lock-and-Key Theory: Exact complementary fitting of the ligand to the receptor at the binding site in a manner similar to that of the key and lock required for the initiation of the response postulated by Emil Fischer in 1894.

Molecular Docking: Prediction of binding mode(s) of a ligand to the binding site of a target biomacromolecule.

Pose: A candidate binding mode of ligand at the receptor binding site calculated based on the docking experiment.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Scoring Functions: Application of mathematical models used for the prediction of binding affinity between ligand and receptor within the molecular docking procedure.

Search Space: All possible orientations and conformations of a ligand as well as the target binding site (receptor) with respect to each other which can be explored during the docking calculation.

Structure-Based Drug Design: Designing and optimizing a ligand structure based on the known 3D-structure of its target molecule.

ENDNOTE

[1] ADMET is an abbreviation for "absorption, distribution, metabolism, excretion, and toxicity"

This work was previously published in Methods and Algorithms for Molecular Docking-Based Drug Design and Discovery edited by Siavoush Dastmalchi, Maryam Hamzeh-Mivehroud, and Babak Sokouti, pages 1-38, copyright year 2016 by Medical Information Science Reference (an imprint of IGI Global).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 31
Docking Methodologies and Recent Advances

Ashwani Kumar
Guru Jambheshwar University of Science and Technology, India

Ruchika Goyal
Guru Jambheshwar University of Science and Technology, India

Sandeep Jain
Guru Jambheshwar University of Science and Technology, India

ABSTRACT

Docking, a molecular modelling method, has wide applications in identification and optimization in modern drug discovery. This chapter addresses the recent advances in the docking methodologies like fragment docking, covalent docking, inverse docking, post processing, hybrid techniques, homology modeling etc. and its protocol like searching and scoring functions. Advances in scoring functions for e.g. consensus scoring, quantum mechanics methods, clustering and entropy based methods, fingerprinting, etc. are used to overcome the limitations of the commonly used force-field, empirical and knowledge based scoring functions. It will cover crucial necessities and different algorithms of docking and scoring. Further different aspects like protein flexibility, ligand sampling and flexibility, and the performance of scoring function will be discussed.

INTRODUCTION

Molecular docking is an entrenched computational technique which can accurately predict the orientation of a ligand and its binding affinity within the constraints of a receptor binding site (Bachwani & Kumar, 2011). Docking of a ligand into the receptor pocket involves several degrees of freedom. Besides the six degrees of translational, rotational, as well as conformational degrees of freedom, some algorithm only uses the translation and orientation degrees of freedom. By increasing computer's performance and using various new algorithms we can explore the ligand conformational degree of freedoms. There are

DOI: 10.4018/978-1-5225-0549-5.ch031

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

simple methods for docking of rigid ligand with rigid receptors and flexible ligand with rigid receptors, but the general method of docking conformations of flexible ligand and receptors is problematic (Leach, Shoichet & Peishoff, 2006).

Besides protein ligand docking, there are protein-protein docking, ligand-nucleic acid docking and protein-nucleic acid docking. Protein-protein docking is important for knowing the structural and functional features of the cell. It is similar to protein ligand docking; scoring and searching algorithms are used. The softwares available for this kind of docking are ZDOCK, HADDOCK, GRAMM-X, Hex protein docking etc. (Ehrlich & Rebecca, 2001). Ligand-nucleic acid docking involves the docking of ligand with nucleic acid like DNA and RNA. Less attention has been made in regard to docking with nucleic acid. Cisplatin, actinomycin and netropsin are example of drugs that shows ligand-DNA docking but less effort has been done in designing of ligands that binds to RNA targets (Mohan, Gibbs, Cummings, Jaeger, & DesJarlais, 2005). Protein-nucleic acid docking requires a set of well defined test cases that forms a common ground for development, validation and comparison of docking method. Software that carried out this kind of docking is HADDOCK (Dijk and Bonvin, 2010).

Earlier 'soft-potential docking' and rotamer libraries were used for flexible receptor docking methods (B-Rao, Subramanian, & Sharma, 2009). Currently X-ray crystallography, NMR, Monte Carlo sampling, Normal Modes-based methods and Molecular Dynamics methods are being focused on constructing an ensemble of structures produced in flexible type docking (Feixas, Lindert, Sinko, & McCammon, 2014; Sousa et al., 2013; Durrant, & McCammon, 2010; Cavasotto, Orry, & Abagyan, 2005). Principle techniques for docking methods use various searching and scoring algorithms like molecular dynamics, Monte Carlo methods, genetic algorithms, fragment-based methods (Mohan, Gibbs, Cummings, Jaeger, & DesJarlais, 2005), point complementary methods, distance geometry methods, tabu searches, evolutionary algorithm, particle swarm optimization and ant colony optimization which are mentioned in detail in later part of this chapter (Taylor, Jewsbury, & Essex, 2002).

FUNDAMENTAL NECESSITIES

Molecular docking program emphasize on the following basic requirements (Mahajan, & Gill, 2014; Krovat, Steindl, & Langer, 2005):

1. A target protein structure with or without a bound ligand is detected by various experimental techniques like NMR or X-Ray crystallography, but if protein structure is not present then protein prediction is done by any technique like threading modelling, homology modelling.
2. Database containing existing or virtual compounds for the docking process
3. Sampling and scoring method, desired scoring and searching algorithms require a computational framework for its efficient working
4. The three-dimensional structure of the protein ligand complex has to be studied in depth of atomic resolution.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

MOLECULAR DOCKING MODELS

1. **Lock and Key Model:** Emil Fischer, in 1890, proposed this model which explains the functioning of biological systems. Ligand fits into the active site of the macromolecule as depicts in Figure 1.
2. **Induced-Fit Model:** Daniel Koshland (1958) introduced the theory of induced-fit. In this, both ligand and receptor mutually interact with each other through small conformational change until an optimal complex is formed (Figure 2).
3. **Conformation Ensemble Model:** Buyong Ma et al., (2003) gave this theory. In this ligand flexibility is achieved by rigid docking of an ensemble of pregenerated ligand conformations with other docking programs.

DOCKING METHODOLOGIES

Docking has been divided in three parts based on the order of increasing complexity (Sousa et al., 2013; Meng, Zhang, Mezei, & Cui, 2011; Mohan et al., 2005):

1. **Rigid Body Docking:** Various docking suites use this type of docking method as first step. In this both ligand and receptor are considered as rigid, which limits the search space. It excludes any kind of flexibility and considers only three types of rotational and three translational degrees of freedom. Examples of software used for this type of docking are R-Dock, Z-Dock, MultiDock, FTDock.
2. **Flexible Ligand Docking/Quasi Docking:** In this type of docking the receptor is kept rigid and ligand is considered as flexible. Gold, FlexX, Glide, AutoDock, ICM, Confort, Omega, Catalyst, Fred are the types of software using this method.
3. **Flexible Docking:** Both receptor and ligand are treated as flexible. Receptor flexibility is still a challenge in this type of docking. It comes under the category of Induced-fit paradigm. In this, both ligand and protein changes their conformations to set into minimum energy fit complex. Also, the high computational cost hurdles its use in screening of large chemical database. Affinity, FlexX-Ensemble, Forecaster and Slide are the software for this category.

Figure 1. Lock and key model

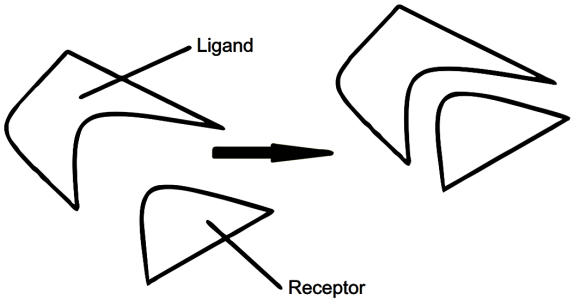
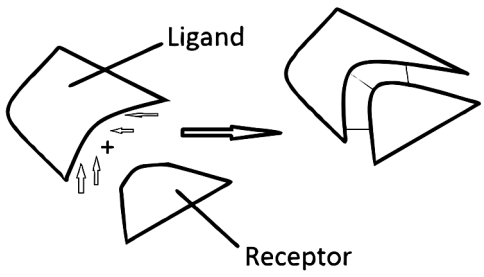


Figure 2. Induced fit model



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

OTHER METHODOLOGICAL ADVANCES

- **Improving Pose Prediction:** Docking algorithms produce a number of docked poses and ranked them based on scoring function, but the dependence only on scoring function sometimes results in false data. So, there is a need of some advances in methodology of scoring function which can improve the prediction of poses for better ranking (Yuriev, Agostino, & Ramsland, 2011).
- **Fragment Based Docking Method:** This method will be discussed in algorithm section of this chapter.
- **Inverse Docking:** Small molecule of interest (ligand) is docked into a library of receptor. It is used for identification and validation of potential targets of various active compounds. This method helps in prediction of various undesirable side-effects like toxicity (Yuriev, & Ramsland, 2013).
- **Blind Docking:** Blind docking is carried out when the ligand binding site is unknown. No major docking software are available that carried out this kind of docking. AutoDock is one of the software that carried out blind docking if structure of ligand, receptor is known without defining *a priori* ligand binding site (Grosdidier, Zoete, & Michielin, 2009)
- **Covalent Docking:** Covalent docking is a tool to study the covalent interactions between ligand and the receptor. It has much more stronger binding affinity that prolongs the duration of biological effect. Amoxicillin, Cefaclor, Isoniazid, Azacitidine, PenicillinV, Warfarin, Orlistat and various other drugs form covalent bond with the target proteins (Kumalo, Bhakat, & Soliman, 2015; Singh et al., 2011). CovalentDock package is used for covalent bond forming ligands. This is used for the accurate and large scale accurate covalent virtual screening. Key features of this package is that for the estimation of covalent binding, it has empirically calibrated energy model and it automatically detect and process structural patterns for the determination of prominent covalent linkage. Beside the advantage of its user friendliness, a limitation of this kind of package is that those who lack in experience face difficulties in compiling, installing and in using this software. To solve this issue, nowadays a fully functional web server the CovalenDock cloud is available (Ouyang, Zhou, Ge, Li & Kwoh, 2013).
- **Surflex Dock:** Surflex dock is used in virtual high throughput screening combined with state of the art speed, accuracy and usability. It requires an updated and re-parameterized scoring function called empirical scoring function, a searching engine and negative training data. It combines a unique scoring function with a patented state of the art search engine. This combination gives excellent docking accuracy and superior screening results. It provides accurate scoring, good speed, easier protein preparation, easier ligand preparation, protomol guided docking, ring flexing, and parallelization (Tripos, L.P., 2010).
- **Docking in Combination with Molecular Dynamics:** Docking only considers the binding affinity of a complex while molecular dynamics (MD) works on the complex shape which changes with time. Sometimes, in spite of the good docking score, it fails in MD simulations. So, it is necessary to validate the docking results with molecular dynamics. Docking is used in combination with molecular dynamics to study the dynamic behaviour of the complexes (Chen, 2015). Lafleur et al., 2009, studied the kinase-ligand binding mode determined by fragment-based docking to validate the MD. It is used as a screening tool for development and optimization of new drug delivery system (Yuriev et al., 2011).
- **Post Processing:** Various procedures are applied to improve the accuracy of the docking and virtual screening performance after taking pose generated by the existing docking and scoring

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

protocols. Example of this kind of procedure is rescoring by MM/PBSA (Molecular Mechanics with Poisson-Boltzmann Surface Area) and MM/GBSA (Molecular Mechanics with Generalized Born Surface Area). Combination of rescoring techniques and multiple docked poses leads to better improvement in the docking methodologies (Meng et al., 2011).

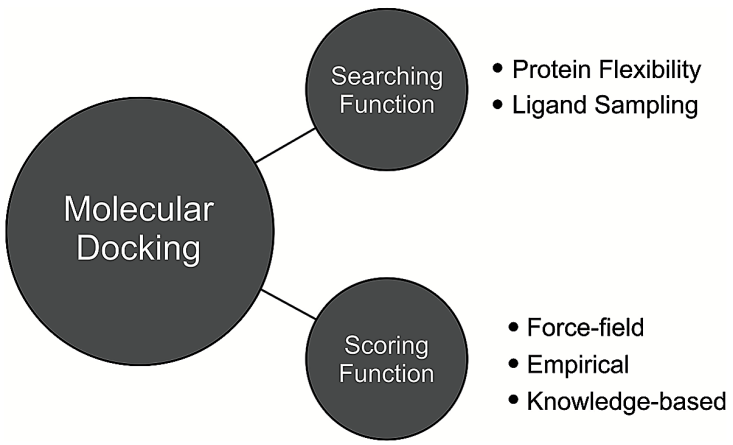
- **Docking into Homology Models:** Docking is a computational calculation based on a range of assumptions and approximations that sometimes leads to some uncertain outcomes. Therefore, if adding another uncertain dimension, that is, a homology model which itself is a result of assumptions and approximations, increases the risk. When experimental data is not easily obtained and the information about complex structures or affinities is required (e.g., for GPCRs), then docking or VS is carried out against homology models and such protocols should be systematically evaluated for their robustness. Docking into various receptor proteins like neuronal nicotinic receptor (Mahasenan et al., 2011), *trypanosomabrucei* phosphodiesterases (Bland et al., 2011), prostaglandin E synthase (He,& Lai, 2011), aryl hydrocarbon receptor transcription factors (Motto, Bordogna, Soshilov, Denison & Bonati, 2011), as well as VS against the s1 receptor (Laurini et al., 2011), calpain (Stuart et al., 2011), kinases (Brylinski, & Skolnick, 2010), and GPCRs (Wada, Kanamori, Nakamura, & Fukunishi, 2011) are the examples of homology models (Yuriev et al., 2013).

As shown in Figure 3, the docking protocol is a combination of two functions: Searching function and scoring function (Hernández-Santoyo, Tenorio-Barajas, Altuzar, Vivano-Cid, & Mendoza-Barrera, 2013; Taylor et al. 2002).

SEARCHING FUNCTION

This function is used to represent the algorithms used for the flexible docking as it considers protein flexibility as well as ligand flexibility. It is used to explore the conformational space of the ligand that binds to the receptor. Various types of docking algorithms will be discussed in later parts of the chapter.

Figure 3. Docking protocol



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Protein Flexibility in Docking

There are many simple methods for docking of rigid ligands with rigid receptors but with flexible receptors is somehow, a formidable challenge in molecular docking studies. It is due to the large size of protein and many degrees of freedom, a small change in receptor conformation leads to increase in the number of receptor conformations (Huang et al., 2010). At the time of docking, a powerful exploration of accessible points for a fully flexible active site is not feasible. Some approaches have been incorporated in order to restrict the number of degrees of freedom like soft docking, side-chain flexibility, molecular relaxation and protein ensemble docking (Huang, & Zou, 2010). In the last few years, advances are made in the field of protein flexibility with the improvement in experimental techniques like NMR, molecular dynamics simulations and sampling techniques. X-Ray crystallography and NMR techniques are basically used for structure determination of proteins (Huang, & Zou, 2007; Legge, Budi, Treutlein, & Yarovsky, 2006). But, some proteins cannot be crystallized then NMR techniques are used for those structures. NMR technique is also used to study the dynamic nature of proteins. But these methods are quite expensive. An alternative of this is taking different conformations snapshot from MD simulations. Some uses Monte Carlo minimization procedure e.g. IFREDA is used for generating ensemble of receptor conformations (Krovat et al., 2005). Cavasotto et al., 2005, gave a new algorithm to generate an ensemble of representative receptor back bone through a combination of normal mode analysis.

Soft Docking

It accounts for only small conformational changes and works by softening the van der Waals interactions and leads to a small degree of overlap between the ligand and receptor. This is computationally efficient method in which only scoring parameters is changed and other parameters remains the same as in case of rigid docking. It can detect only subtle changes (Sousa et al., 2013; Huang et al., 2010).

Side-Chain Flexibility

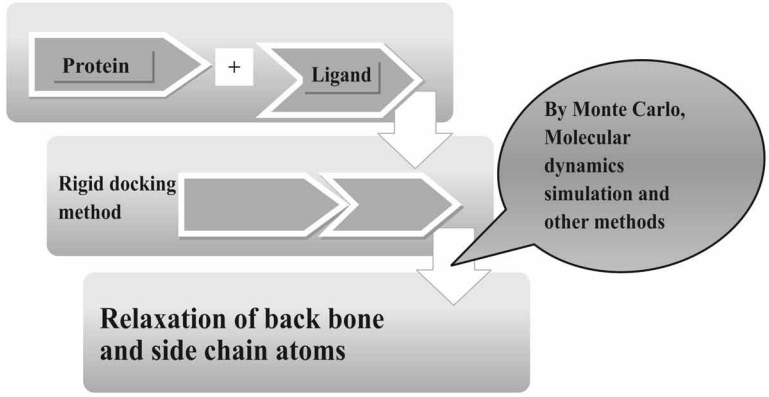
Discrete side-chain flexibility occurs due to the exploration of rotamer libraries. This method fails to fully simulate ligand-induced flexibility as backbone conformations are fixed only side chain conformations are flexible (Sousa et al., 2013; Huang et al., 2010). Zavodsky and Kuhn (2005) proposed a 'minimal rotation hypothesis' and used 'SLIDE' as a docking algorithm. They made several attempts to sort out ligand–receptor steric clashes by a minimizing the number of side-chain conformations (Totrov, & Abagyan, 2008).

Molecular Relaxation

Despite being time consuming, this type of method is more demanding than side-chain flexibility method as it involves both side chain as well as back bone conformations. It first involves the method of rigid docking to fit the ligand into the receptor. After that the various relaxation procedures are used to relax the backbone and side-chain atoms like Monte Carlo, Molecular Dynamics simulations (Huang et al., 2010) and others as illustrated in Figure 4.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 4. Representation of molecular relaxation



Protein Ensemble Docking

It is the most widely used protein flexibility method which utilizes multiple protein structures to represent various possible conformations (Huang et al., 2010). In this method, side-chain and back bone conformation changes in restricted manner. Multiple protein structure generates an average receptor grid by seeing the flexibility of multiple structures which therefore, significantly minimizes the cost. Ensemble of docking is generated by experimental procedure like NMR and others such as Monte Carlo simulations and Molecular Dynamics simulations (Huang et al., 2010).

Ligand Sampling and Flexibility

Ligand sampling is the most prominent area in the field of molecular docking by exploring all possible ligand degrees of freedom to get the experimental binding orientation. It requires sampling algorithms to explore various ligand conformations and their binding with the proteins. Binding site can be active site, a dimer interface or other site of interest (Huang et al., 2010). Ligand sampling is done by two approaches, *viz*, whole molecule approach and fragment based sampling approach. To sample the pose space whole molecule method is used and it produces various pre-calculated ligand conformations by Monte Carlo (MC), Molecular Dynamic simulations (MD), Genetic algorithms (GA), simulated annealing (SA) and fragment-based techniques. A rough idea of three types of sampling algorithms (shape matching, systematic search and stochastic algorithms) is given below:

Shape Matching

It is the simplest method used for ligand sampling. As the name depicts, it places the ligand in the position that its molecular surface is in complement with the molecular surface of the binding site of the receptor. There are six degrees of freedom for the ligand binding orientation. DOCK, FRED, FLOG, EUDOC, LigandFit, Surflex, MS-DOCK, and MDock are the examples for this type of method (Huang et al., 2010).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

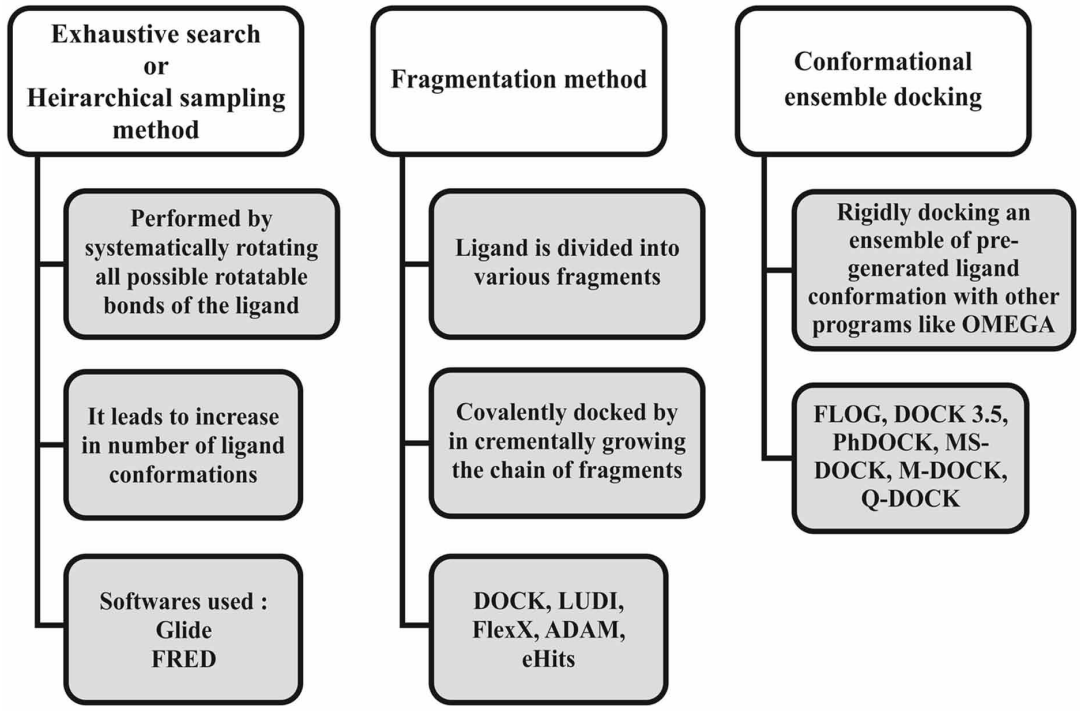
Systematic Searches

This method is commonly used for flexible-ligand docking that generates all possible ligand binding conformations after going through all degrees of freedom of the ligand. Exhaustive search, fragmentation and conformational ensemble methods are the three types of systematic search algorithms. It systematically endures all possible conformations and demonstrates the brute force solution to the docking problem. All molecules are supposed to be rigid and evaluation of their interaction energy from a force field model is done. Glide and FRED are two exhaustive examples of systematic search (Huang et al., 2010). Systematic search methods are of three types as depicted in Figure 5.

Stochastic Algorithms

Sampling of ligand binding orientations and conformations is done by making random changes to either a single ligand or population of ligands (Stark, & Powers, 2012). Four popular random approaches are Monte Carlo (MC) methods, evolutionary algorithms, Tabu search methods, Particle swarm optimization method, Ant colony optimization (Huang et al., 2010).

Figure 5. Types of systematic search methods



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

SCORING FUNCTIONS

Implication of an acceptable solution to the sampling problem is very challenging in finding out the correct binding mode. The rationale behind the use of the scoring procedure is to identify the correct binding pose by its lowest energy value and also to rank the protein-ligand complexes in accordance with their binding affinity which is calculated as a summation of independent terms. Scoring function depends upon the size of the molecule which is in proportionality with score, its entropy, and the solvation and desolvation effects. Scoring functions are subdivided into force-field, empirical, and knowledge based approaches (Schulz-Gasch, & Stahl, 2004).

Force-field based approach is used in protein-ligand docking as well as in molecular dynamics simulations. Binding affinity is calculated by adding van der Waals interactions, electrostatic interactions, bond stretching/ bending and torsional forces. The two shortcomings of this method are steric clash that cannot be avoided and the exaggerating polar interactions which can be compensated by using desolvation energy. AMBER and CHARMM (Wong, 2008) are the two examples of this category. DOCK is the typical force-field scoring function whose energy parameters are taken from AMBER force fields.

$$E = \sum_i \sum_j \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{\in (r_{ij}) r_{ij}} \right)$$

where, r_{ij} stands for the distance between protein atom i and ligand atom j, A_{ij} and B_{ij} are the VDW parameters, and q_i and q_j are the atomic charges. ε (r_{ij}) is usually set to 4r, reflecting the screening effect of water on electrostatic interactions. The effect of solvent is determined by using distance-dependent dielectric constant but it does not account for desolvation effect (Huang et al., 2010).

Empirical based approach is based on hydrophilic and hydrophobic contacts. In spiteof the key issues like the proper modelling of repulsive forces as well as the correct balance between hydrophilic and hydrophobic contacts, it is very successful in modeling complex molecular recognition phenomena with a very simplistic interaction model. Examples include LUDI (Wong, 2008), ChemScore, and X-SCORE.

Estimation of binding affinity of a complex is based on a set of weighed energy terms.

$$\triangle G = \sum_i W_i \, . \triangle G_i$$

where, ΔG_i represents individual empirical energy terms such as VDW energy, electrostatics, hydrogen bond, desolvation, entropy, hydrophobicity, etc. (Huang et al., 2010). The corresponding coefficients $W_i$are determined by reproducing the binding affinity data of training set of protein-ligand complexes with known three-dimensional structures, using least squares fitting.

It is faster than force field in binding energy calculations as it uses the simple energy terms.

Knowledge based approach is also known as statistical potential-based scoring function. As the name indicates, it is based on statistical analysis of experimentally determined complex structures. In this, the potential parameters are derived based on the structural information of the protein-ligand complex and after that sum of interactions of all protein-ligand atom pairs present in the complex are taken. The method shows a good balance between speed and accuracy as compared to the above two mentioned

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

approaches. Some problems persist with this approach like excluded volume, reference state and sparse data problem. DFIRE, DrugScore, ITScore, PMF-score, and SMOG (Wong, 2008) are some of the examples based on the knowledge based scoring functions. The potential is calculated by using inverse Boltzmann relation:

$$w(r) = -\kappa_B T \ln\left[\frac{\rho(r)}{\rho^*(r)}\right]$$

where, κ_B is the Boltzmann constant, T is the absolute temperature of the system, $\rho(r)$ is the number density of the protein–ligand atom pair at distance r, and $\rho^*(r)$ is the pair density in a reference state where the inter-atomic interactions are zero (Huang et al., 2010).

OTHER ADVANCEMENT IN SCORING FUNCTION

Quantum Mechanics Method (QM Method)

Nowadays, Quantum mechanics method is boomed in the area of computational drug design and development due to its high accuracy for the estimation of binding affinities. For low to medium throughput *in-silico* screening, linear scaling QM was used for efficient approximations. QM based procedures are being used for the estimation of protonation states of ionisable groups, evaluating energies and molecular structure optimization. In HTS, it is employed to derive robust quantitative structure activity relationship (Zhou, Huang, & Caflisch, 2010). Raha and Merz gave this approach to overcome the inability of the above three approaches for a set of 18 carbonic anhydrase and five carboxy peptidase inhibitors (Krovat et al., 2005).

Consensus Scoring

Recent advances are made by researchers to increase the performance of scoring functions in spite of their limitations. This can be achieved by combining several scoring functions by using consensus scoring (Yuriev, Agostino, & Ramsland, 2011). It is used to lessen the false positive results of virtual screening and also to improve the structure prediction by improving the probability of finding a correct solution in hope to balance out the errors of the individual scoring functions. Examples of consensus scoring are MultiScore and X-Cscore (Huang et al., 2010).

Clustering and Entropy Based Scoring Methods

This method is also an advancement in improving the performance of scoring function which includes the entropic effects by dividing generated ligand binding modes into different clusters. The contribution of entropy is measured by the configurational space covered by the number of ligand poses in the cluster. Limitation of this method is that this approach is docking program dependent i.e. its performance depends upon the ligand sampling protocol used. If this method is used in combination with AutoDock, it improves the binding mode prediction (Yuriev, Agostino, & Ramsland, 2011).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Fingerprinting

This method identifies various interaction patterns that can guide and improve predictions. Interaction fingerprints (Ifs) and profile based methods have been used in scoring and virtual screening (Yuriev, Agostino, & Ramsland, 2011). For example Glide XP descriptors for locating key pharmacophoric features of docked fragments (Perez-Nueno, Rabal, Borrell, & Teixido, 2009) and weighted-Structural Interaction Fingerprints (w-SIFt) are used to determine inhibitor potency (Nandigam, Kim, Singh, & Chuaqui, 2013).

Combination with Experimental Data

Gonzalez-Ruiz, & Gohlke, 2009, combined drug score function with experimental data, specifically NMR amide proton chemical shift perturbations (CSP) which lead to improvement in pose prediction.

Target Specific Scoring Functions

It leads to improved robustness of the scoring function parameters that refrain to multiple targets of a target class (Yuriev et al., 2013). Target specific target system were developed for Kinase-PMF to assess the binding of ATP-competitive kinase inhibitors, HIV protease, carbonic anhydrase, and trypsin; enoyl acyl carrier protein reductase from Mycobacterium tuberculosis (Seifert, 2009).

PERFORMANCE OF SCORING FUNCTION

The three related and independent criteria are binding mode identification, binding affinity prediction and virtual screening database which are used to assess the performance of a scoring function. Firstly, the determination of binding mode is one of the essential features for the scoring function. The structure is ranked by their calculated binding scores. The rmsd value should be ≤ 2.0 Å between the top ligand conformation and the experimental value for the best result. However, this theory sometimes fails in case of nearly symmetrical ligands, large flexible ligand, solvent group and unimportant group which sometimes hide the performance of docking program. To solve this issue various other parameters like relative displacement error (RDE), interaction-based accuracy classification (IBAC), real space R-factor (RSR), and Generally Applicable Replacement for rmsD (GARD) are used. The tightness of the ligand binding to the receptor represents the binding affinity. It is difficult to achieve a score scale related to experimental binding data. To determine the correlation between theoretical and experimental data, the Pearson correlation coefficient and Spearman correlation coefficient are calculated. Prediction of the binding affinity is more challenging than binding mode because of various experimental errors and different experimental conditions. To assess the performance of scoring function, the enrichment test and the area under the receiver operating characteristic curve (AUC) of virtual data base screening are done. For a perfect scoring function, it is necessary to perform equally well on all the three criteria of any test set. However, due to the innate restrictions, most of the existing scoring functions executes well with only one or two of the criteria and fail on others (Huang, Grinter, & Zou, 2010).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

DIFFERENT ALGORITHMS USED FOR DOCKING

Docking algorithms are used to obtain a fast and accurate method for the discovery of various novel lead compounds or to validate the experimental data (Dias, & de Azevedo, 2008). Here are some of the different algorithms used for molecular docking studies:

- **Molecular Dynamics:** Molecular dynamics (MD) is extensively used as a powerful simulation technique in the field of molecular modeling. The calculation of this method is based on Newton's equation of motions. It depicts the flexibility of both the ligand and protein more efficiently than any other algorithms. However, the disadvantage of MD simulations is that they progress petitely and thus have difficulties in reaching global energy minimum, which may results in improper sampling (Taylor et al., 2002). It is a very important technique as docking only considers binding affinity but MD accentuates the conformational change of the complex with time, thus, reducing the calculation capacity (Ian et al., 2013). So, MD simulation is considered as a validation before concluding any result of docking.
- **Monte-Carlo Simulation Method:** Monte Carlo (MC) methods produce various ligand poses during bond rotation, rigid-body translation or rotation. MC method is superior to MD to explore the search space by crossing the high energy barrier in potential energy space. It searches ligand conformational pose by random torsional sampling of rotatable bonds. Glide, ICM and Rosetta ligand methods are used to generate the pre-calculated ligand (Huang, 2014).
- **Genetic Algorithm Method:** Genetic algorithm (GA) is a well-known stochastic method. Idea of its stem comes from Darwin's theory of evolution. The ligand translational, rotational, and torsional degrees of freedom are considered for this type of algorithm. Genes are encoded as binary strings for degrees of freedom that make up the'chromosome' which represents the pose of the ligand. The genetic operators i.e. mutation and crossover affects the gene which results in a new ligand structure. Gold, Divali, Darwin and AutoDock programs are the softwares employed for genetic algorithms (Taylor et al., 2002).
- **Fragment Based Method:** It dissects the ligand molecule into a base fragment and various flexible fragments. Firstly, base fragment is docked in optimal orientation to establish favourable complementary interactions. After that, the addition of flexible fragments in a stepwise manner to that protein, explores all possible bond rotational conformations. This method can further be extended to de novo design programs (Taylor et al., 2002). FlexX, Hammerhead, and DOCK programs use fragment based ligand sampling methods (Chen, & Pohlhaus, 2010).
- **Point Complementary Methods:** Ligand can be docked into the binding site of the protein using point complementary method.This method is based upon the shape and chemical similarity between the interacting molecules. Perhaps, fragment based docking algorithms could also be included in this category, for that a division should be made between algorithms used for point complimentary and fragment based method. Examples of programs using point complementary methods are FTDOCK, SANDOCK, FLOG and the soft docking algorithm (Taylor et al., 2002).
- **Distance Geometry Methods:** It determines the binding mode between ligand and protein and considers only hydrogen bonding. It enables fast sampling of conformational space and thus helps in identifying palusible binding modes and is often used to express embedding algorithm. A recent example of a program using distance geometry in docking is DockIt. The resulting structures are then scored using PMF or PLP scoring functions (Taylor et al., 2002).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

- **Tabu Searches:** This method is based on stochastic process in which n numbers of small random changes are made to the current conformation; the new solutions obtained are then scored and ranked in ascending order. Accept the best modification having a lower value than any other accepted modification even it is in the tabu list or not (Kitchen, Decornez, Furr, & Bajorath, 2004). An example of docking algorithm using tabu search is PRO LEADS (Taylor et al., 2002).
- **Evolutionary Algorithms:** This is the most commonly used stochastic method for flexible ligand docking. It is used to sample most favourable energetic conformation among the large search space of protein-ligand complexes (Thomsen R. 2003).
- **Ant Colony Optimization (ACO):** This method of optimization is inspired by the behaviour of real ants. Deposition of pheromone enables real ants to find the shortest route between the food and their nests without visual information. Similarly, in this method artificial ants can be deployed to solve complex optimization problems by using artificial pheromone deposition. Protein-Ligand ANT System is based on this type of algorithm used in protein-ligand docking. The artificial ants are used to obtain a minimum energy conformation of the ligand in the binding site with pheromone trails. This software is highly used for virtual screening of a large number of databases and lead optimization (Kumar et al., 2014). This software can be used for rigid body docking of multiple conformations of ligands into rigid as well as flexible receptors (Korb, Stutzle, & Exner, 2006).
- **Particle Swarm Optimization Method (PSO Method):** Dr. Eberhart and Dr. Kennedy, 1995, developed a population based stochastic optimization technique. They were inspired by the social behaviour of flying birds. As the name swarm depicts the behaviour of a large group of insects or birds, in this method a particle fly in the search space by following the current optimum particles. Movement of a ligand through search space is based on the concept of changing velocity of each particle towards its *p*best and *l*best locations.

PSO is based on the behaviour of birds flocking. Suppose a group of birds in the sky are searching for food in an area and there is only one food in that area. Here, the birds are called as solutions or particle and food is optima by updating generations. All birds don't know the exact location of food but they know how far the food is from them in each iteration. So, to find the food they follow the bird which is nearest to the food. PSO used this phenomenon to solve the optimization problems by following the current optimum particles. All particles are having their optimized fitness function for the evaluation of fitness value and velocities to direct the flying of the particles. In every iteration, each particle is updated by following two best values:

1. The best solution, which is achieved, gives best fitness value is called as *p*best
2. Another best value which is tracked with PSO by any particle in the population is called as global best (*g*best). *l*best when particle takes all the particles as its topological neighbours.

After calculating the two best values, calculation of its velocity and position is done with the following equations (a) and (b).

$$v[\ ] = v[\ ] + c1 * rand(\ ) * \left(pbest[\ ] - present[\ ]\right) + c2 * rand(\ ) * \left(gbest[\ ] - present[\ ]\right) \quad \text{(a)}$$

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

$$present[\ ] = present[\ ] + v[\ ] \quad \text{(b)}$$

v $v[\ ]$ is the particle velocity, present [ ] is the current particle (solution). *p*best [ ] and *g*best [ ] are already defined, rand () is a random number between (0,1). c1, c2 are learning factors, usually c1 = c2 = 2.

Comparison of PSO with GA

PSO is similar to GA as it involves a population of random solutions and searches for optima by updating generations but no evolution operators such as crossover and mutation are involved. PSO can be used for a wide range of applications or to specific application determined for a specific requirement. It is a faster and cheaper technique as compared to other methods. Only few parameters need to be adjusted in PSO and that are easy to implement (Hu, Eberhart, & Shi, 2004).

A GLIMPSE OF VIRTUAL SCREENING

Structural based virtual screening (VS) begins with the use of computing system to identify ligand-binding site on the target molecule from a large database. It is a complementary approach to experimental screening methods. There are two types of VS, one is structure-based virtual screening (docking) and other is ligand-based virtual screening (Reddy, Pati, Kumar, Pradeep, & Sastry, 2007). The basic goal of VS is to decrease the cost of experimental screening and to identify various lead compounds. The whole process depends upon information available e.g. if 3D-structure is available then molecular docking method is used or if 3D-structure is not available then pharmacophore model is derived based upon various properties like molecular weight, lipophilicity, ADME properties. Recent advances have been done in the field of virtual screening based on molecular docking studies like improvement of old docking programs for example DOCK, FlexX (Tripos, 2005), AutoDock, Gold (Cambridge Crystallographic Data Centre, 1965-2015), ICM (MolSof, 2015). Several new docking programs which have been released in past decade, including EUDOC, LigandFit and eHiTS are used for VS (Hou, & Xu, 2004).

APPLICATIONS OF MOLECULAR DOCKING

Several research groups focus on evaluating the performance of various docking programs or on making improvements to the scoring functions when experimental testing has already been done. Such efforts could give meaningful guidance to choose the methodology for a particular target system and has a wide variety of applications. Docking and VS both are important for the identification of various hits, lead optimization for binding pose prediction, bioremediation. It results in a number of success stories, it is not a separate technique but it is entrenched in between the *in-silico* and experimental technique (Eweas, Maghrabi, & Namarneh, 2014). Some novel docking softwares along with their features and scoring details are given in Table 1.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 1. Some novel docking programs along with their key features and scoring functions

Programs	Key Features	Scoring Function
AutoDock Vina	• Automation of input and output • No limit on variables (number of atoms, rotatable bonds,grid map size, etc.) • Parallelism and multithreading on multicore machines	Vina (ML)
BetaDock	• Prioritizing shape complementarity, based on the theory of b-complex and the Voronoi diagram • Rigid bodies (both receptor and ligand)	AMBER + AutoDock (FF)
bhDock	• Identification of low-resolution binding sites • Ligand posing via simulated annealing MD-based global optimization	
EADock DSS	Efficient tree-based dihedral space sampling (DSS), based on hybrid sampling engine and multiobjective scoring	CHARMM22 (FF)
H-DOCK	Docking by hydrogen bond matching and shape complementarity	H-DOCK (HBM and SSC terms)
LigDockCSA	• Powerful global optimizer • Conformational space annealing (CSA)	AutoDock (FF) + PLP torsional energy term
MedusaDock	Modeling both ligand and receptor flexibility simultaneously with sets of discrete rotamers	MedusaScore (FF)
PD-DOCK	Parameter optimization for docking scores	PD-DOCK (K)

CASE STUDY

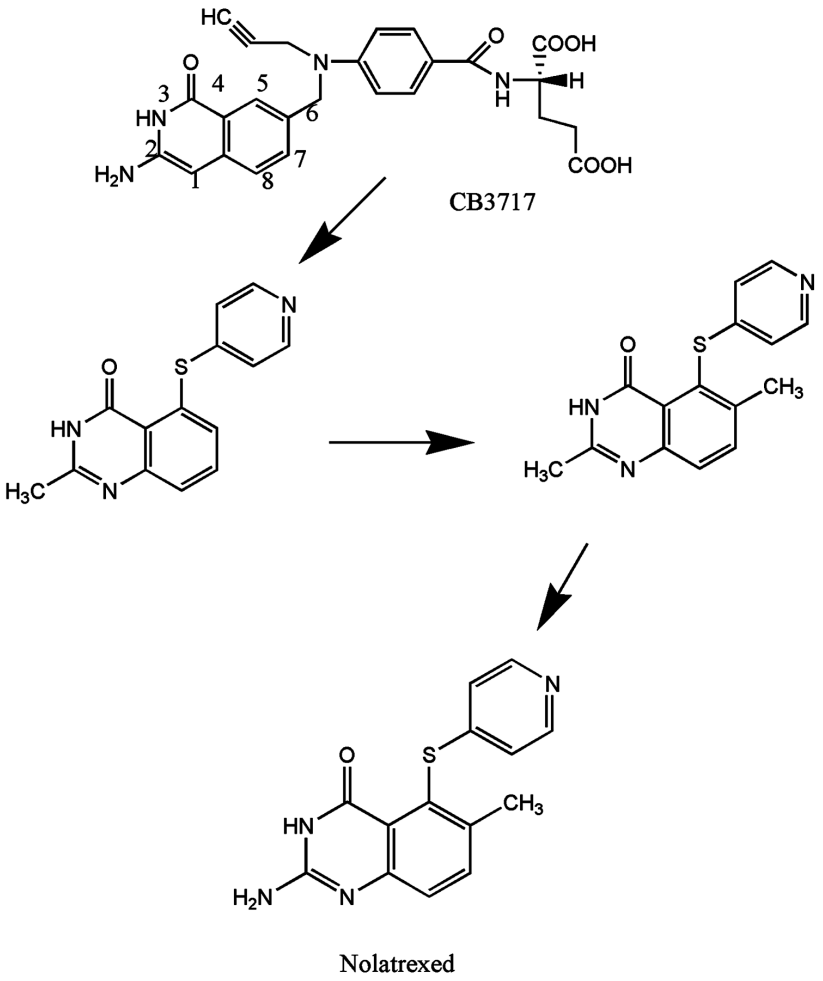
Nolatrexed used for the treatment of unresectable hepatocellular carcinoma (HCC) found to have high inhibitory potential of thymidylate synthase (TS). Its discovery involves different structure-based molecular modelling methods and X-ray crystal structure data on TS of *Escherichia coli* which is highly homologous to human TS which led to the design of a potent quinazoline derivative CB3717. Various modifications in the compound CB3717 were carried out at position 2, 5, and 6 of ring and some were kept constant like 4-ketoquinazoline ring that accounted for hydrogen bonding and pi_pi stacking interaction and gave compound Nolatrexed (Figure 6). It completed the Phase III clinical trial (Roa, Kar, & Das, 2015).

Boceprevir is used for the treatment of hepatitis C. It inhibits NS3-NS4A serine protease in hepatitis C virus (HCV), leading to the prevention of the replication of virus particles. Earlier high-throughput screening (HTS) analysis attempts for the development of HCV failed, that led to a structure-based design approach for potential molecules. Initially, it was observed that the inhibition of NS3-NS4A protease is facilitated by compounds that release N-terminal peptides. Later, it was found that Ser 139 in the viral protease was a target for electrophilic reaction with groups like aldehyde and ketones. Schering Plough Research Institute designed a series of α-ketoamide inhibitors. This drug was approved by FDA in 2011 (Roa, Kar, & Das, 2015).

Raltegravir, is a well-known drug approved by FDA in 2007 for the treatment of AIDS. It is a prominent HIV-1 integrase inhibitor. It is the first FDA-approved inhibitor. Binding site of this receptor was used in the molecular dynamics trajectory. Software used for the discovery was AutoDock along with the relaxed complex method. Initial molecules that showed inhibitory potential of HIV-integrase-catalyzed transfer of DNA were two classes of chemicals (namely 5, 6-dihydroxypyrimidine-4-carboxamides and N-methyl-4-hydroxypyrimidinone-carboxamides) and then through a series of structural changes and optimization, Raltegravir was discovered (Roa, Kar, & Das, 2015) as shown in Figure 7.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 6. Case study of Nolatrexed



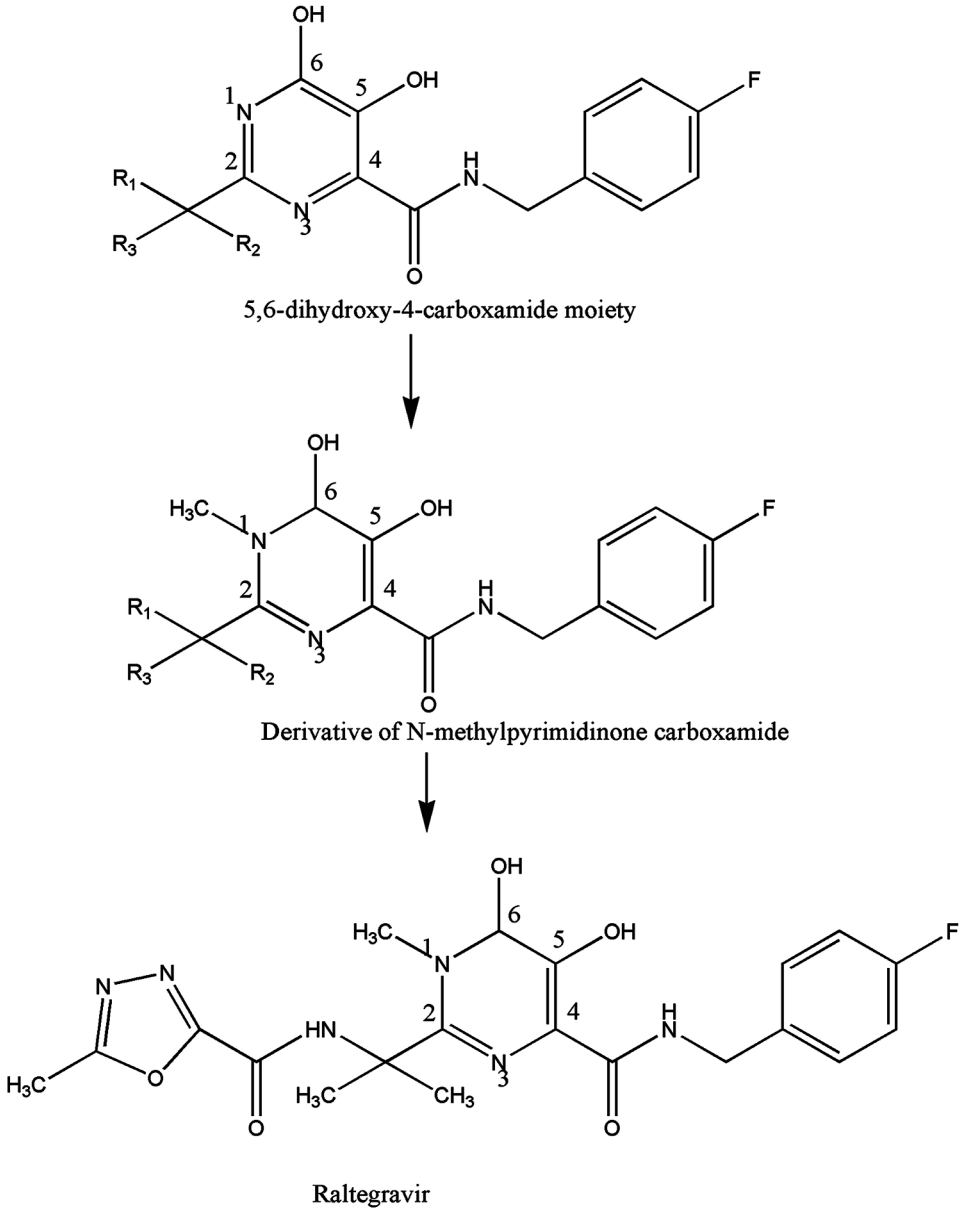
Nolatrexed

CONCLUSION

Molecular docking exhibit potential effect on identification and optimization in modern drug discovery. It helps us to rank the various lead compounds, predict the binding sites and affinities and elucidate the binding mechanism. Docking based virtual screening leads to much higher hit rate than traditional screening methods (HTS). It is not the replacement of experimental methods but both docking and experimental screening methods are complimentary and will reduce the cost of drug discovery to a large extent. Hopefully, innovative advancement in this field is expected in future which will ease the process of research on novel therapeutically active moieties.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 7. Case study of Raltegravir



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

REFERENCES

Abid, O. U. R., Babar, T. M., Ali, F. I., Ahmed, S., Wadood, A., Rama, N. M., & Choudhary, M. I. et al. (2010). Identification of novel urease inhibitors by high-throughput virtual and *in-vitro* screening. *ACS Medicinal Chemistry Letters*, *1*(4), 145–149. doi:10.1021/ml100068u PMID:24900188

Allenby, G., Bocquel, M.-T., Saunders, M., Kazmer, S., Speck, J., Rosenberger, M., & Chambon, P. et al. (1993). Retinoic acid receptors and retinoid X receptors: Interactions with endogenous retinoic acids. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(1), 30–34. doi:10.1073/pnas.90.1.30 PMID:8380496

Bachwani, M., & Kumar, R. (2011). Molecular docking: A Review. *International Journal of Research in Ayurveda and Pharmacy*, *2*(6), 1746–1751.

Bailey, D., & Brown, D. (2001). High-throughput chemistry and structure-based design: Survival of the smartest. *Drug Discovery Today*, *6*(2), 57–59. doi:10.1016/S1359-6446(00)01596-8 PMID:11166243

Barril, X., & Fradera, X. (2006). Incorporating protein flexibility into docking and structure-based drug design. *Expert Opinion Drug Discovery*, *1*(4), 1–15. doi:10.1517/17460441.1.4.335 PMID:23495904

Bland, N. D., Wang, C., Tallman, C., Gustafson, A. E., Wang, Z., Ashton, T. D., & Pollastri, M. P. et al. (2011). Pharmacological validation of Trypanosoma brucei phosphodiesterases B1 and B2 as druggable targets for African sleeping sickness. *Journal of Medicinal Chemistry*, *54*(23), 8188–8194. doi:10.1021/jm201148s PMID:22023548

Borkotoky, S. (2012). Docking studies on HIV integrase inhibitors based on potential ligand binding sites. *International Journal of Bioscience*. *Biochemistry and Bioinformatics*, *2*(3), 21–29.

Brylinski, M., & Skolnick, J. (2010). Comprehensive structural and functional characterization of the human kinome by protein structure modeling and ligand virtual screening. *Journal of Chemical Information and Modeling*, *50*(10), 1839–1854. doi:10.1021/ci100235n PMID:20853887

Bustanji, Y., Al-Masri, I. M., Qasem, A., Al-Bakri, A. G., & Taha, M. O. (2009). *In silico* screening for non-nucleoside HIV-1 reverse transcriptase inhibitors using physicochemical filters and high-throughput docking followed by *in vitro* evaluation. *Chemical Biology & Drug Design*, *74*(3), 258–265. doi:10.1111/j.1747-0285.2009.00852.x PMID:19703027

Buzko, O. V., Bishop, A. C., & Shokat, K. M. (2002). Modified AutoDock for accurate docking of protein kinase inhibitors. *Journal of Computer-Aided Molecular Design*, *16*(2), 113–127. doi:10.1023/A:1016366013656 PMID:12188021

Capdeville, R., Buchdunger, E., Zimmermann, J., & Matter, A. (2002). Glivec (STI571, imatinib), a rationally developed, targeted anticancer drug. *Nature Reviews. Drug Discovery*, *1*(7), 493–502. doi:10.1038/nrd839 PMID:12120256

Carlson, H. A. (2002). Protein flexibility and drug design: How to hit a moving target. *Current Opinion in Chemical Biology*, *6*(4), 447–452. doi:10.1016/S1367-5931(02)00341-1 PMID:12133719

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Cavasatto, C. N., Kovacs, J. A., & Abagyan, R. A. (2005). Representing receptor flexibility in ligand docking through relevant normal modes. *Journal of the American Chemical Society*, *127*(26), 9632–9640. doi:10.1021/ja042260c PMID:15984891

Cavasotto, C. N., Orry, A. J. W., & Abagyan, R. A. (2005). The challenge of considering receptor flexibility in ligand docking and virtual screening. *Current Computer-aided Drug Design*, *1*(4), 423–440. doi:10.2174/157340905774330291

Chen, Y. & Pohlhaus, D. T. (2010). In silico docking and scoring of fragments. *Drug Discovery Today: Technologies | Fragment based drug discovery, 7*(3), 149-156.

Chen, Y.-C. (2015). Beware of docking! *Trends in Pharmacological Sciences*, *36*(2), 78–95. doi:10.1016/j.tips.2014.12.001 PMID:25543280

Cohen, P., & Alessi, D. R. (2013). Kinase drug discovery– What's next in the field? *ACS Chemical Biology*, *18*(8), 96–104. doi:10.1021/cb300610s PMID:23276252

DesJarlais, R. L., Seibel, G. L., Kuntz, I. D., Furth, P. S., Alvarez, J. C., Babe, L. M., & Craik, Y., C. S. (1990). Structure-based design of nonpeptide inhibitors specific for the Human Immunodeficiency Virus 1 protease.*Proceedings of the National Academy of Sciences*, *87*,6644-6448. doi:10.1073/pnas.87.17.6644

Dias, R., & de Azevedo, W. F. Jr. (2008). Molecular Docking Algorithms. *Current Drug Targets*, *9*(12), 1040–1047. doi:10.2174/138945008786949432 PMID:19128213

Dijk, M. V., & Bonvin, A. M. J. J. (2010). Pushing the limits of what is achievable in protein–DNA docking: Benchmarking HADDOCK's performance. *Nucleic Acids Research*, *38*(17), 5634–5647. doi:10.1093/nar/gkq222 PMID:20466807

Durrant, J. D., & McCammon, J. A. (2010). Computer-aided drug-discovery techniques that account for receptor flexibility. *Current Opinion in Pharmacology*, *10*(6), 77–0774. doi:10.1016/j.coph.2010.09.001 PMID:20888294

Ehrlich, L. P., & Rebecca, C. (2001). Protein-protein docking. *Reviews in Computational Chemistry*, *17*, 61–98. doi:10.1002/0471224413.ch2

Eweas, A. F., Maghrabi, I. A., & Namarneh, A. I. (2014). Advances in molecular modeling and docking as a tool for modern drug discovery. *Der Pharma Chemica*, *6*(6), 211–228.

Fabbro, D. (2015). 25 years of small molecular weight kinase inhibitors: Potentials and limitations. *Molecular Pharmacology*, *87*(5), 766–775. doi:10.1124/mol.114.095489 PMID:25549667

Feixas, F., Lindert, S., Sinko, W., & McCammon, J. A. (2014). Exploring the role of receptor flexibility in structure-based drug discovery. *Biophysical Chemistry*, *186*, 31–45. doi:10.1016/j.bpc.2013.10.007 PMID:24332165

Fraczek, T., Siwek, A., & Paneth, P. (2013). Assessing molecular docking tools for relative biological activity prediction: A case study of triazole HIV-1 NNRTIs. *Journal of Chemical Information and Modeling*, *53*(12), 3326–3342. doi:10.1021/ci400427a PMID:24266618

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Gane, P. J., & Dean, P. M. (2000). Recent advances in structure-based rational drug design. *Current Opinion in Structural Biology*, *10*(4), 401–404. doi:10.1016/S0959-440X(00)00105-6 PMID:10981625

Ghanemi, A. (2015). Targeting G protein coupled receptor-related pathways as emerging molecular therapies. *Saudi Pharmaceutical Journal*, *23*(2), 115–129. doi:10.1016/j.jsps.2013.07.007 PMID:25972730

Gonzalez-Ruiz, D., & Gohlke, H. (2009). Steering protein–ligand dockingwith quantitative NMR chemical shift perturbations. *Journal of Chemical Information and Modeling*, *49*(10), 2260–2271. doi:10.1021/ci900188r PMID:19795907

Grinter, S. Z., & Zou, X. (2014). Challenges, Applications, and Recent Advances of Protein-Ligand Docking in Structure-Based Drug Design. *Molecules (Basel, Switzerland)*, *19*(7), 10150–10176. doi:10.3390/molecules190710150 PMID:25019558

Grosdidier, A., Zoete, V., & Michielin, O. (2009). Blind docking of 260 protein-ligand complexes with EADock 2.0. *Journal of Computational Chemistry*, *00*, 1–10. PMID:19130502

Harris, R. A. (2013). Functional Validation of Virtual Screening for Novel Agents with General Anesthetic Action at Ligand-Gated Ion Channels. *Molecular Pharmacology*, *84*(5), 670–678. doi:10.1124/mol.113.087692 PMID:23950219

He, S., & Lai, L. (2011). Molecular docking and competitive binding studydiscovered different binding modes of microsomal prostaglandin E synthase-1 inhibitors. *Journal of Chemical Information and Modeling*, *51*(12), 3254–3261. doi:10.1021/ci200427k PMID:22077876

Hernández-Santoyo, A., Tenorio-Barajas, A. Y., Altuzar, V., Vivano-Cid, H. & Mendoza-Barrera, C. (2013). *Protein-Protein and Protein-Ligand Docking,* Croatia, Europe: INTECH

Heusser, S. A., Howard, R. J., Borghese, C. M., Cullins, M. A., Carlsson, J., Hou, T., & Xu, X. (2004). Recent Development and Application of Virtual Screening in Drug Discovery: An Overview. *Current Pharmaceutical Design*, *10*(9), 1011–1033. doi:10.2174/1381612043452721 PMID:15078130

Hoi, P. M., Li, S., Vong, C. T., Tseng, H. H. L., Kwan, Y. W., & Lee, S. M.-Y. (2015). Recent advances in structure-based drug design and virtual screening of VEGFR tyrosine kinase inhibitors. *Methods (San Diego, Calif.)*, *71*, 85–91. doi:10.1016/j.ymeth.2014.09.004 PMID:25239735

Hu, X., Eberhart, R., & Shi, Y. (2004). *Recent advances in particle swarm*. IEEE Congress on Evolutionary Computation 2004, Portland, OR.

Huang, S.-Y. (2014). Search strategies and evaluation in protein–protein docking: Principles, advances and challenges. *Drug Discovery Today*, *19*(8), 1081–1096. doi:10.1016/j.drudis.2014.02.005 PMID:24594385

Huang, S.-Y. (in press). Exploring the potential of global protein-protein docking: An overview and critical assessment of current programs for automatic *ab initio* docking. *Drug Discovery Today*. PMID:25801181

Huang, S.-Y., Grinter, S. Z., & Zou, X. (2010). Scoring functions and their evaluation methods for protein–ligand docking: Recent advances and future directions. *Physical Chemistry Chemical Physics*, *12*(40), 12899–12908. doi:10.1039/c0cp00151a PMID:20730182

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Huang, S.-Y., & Zou, X. (2007). Efficient molecular docking of NMR structures: Application to HIV-1 protease. *Protein Science*, *16*(1), 43–51. doi:10.1110/ps.062501507 PMID:17123961

Huang, S.-Y., & Zou, X. (2007). Ensemble Docking of Multiple Protein Structures: Considering Protein Structural Variations in Molecular Docking. *PROTEINS: Structure, Function, and Bioinformatics*, *66*(2), 399–421. doi:10.1002/prot.21214 PMID:17096427

Huang, S.-Y., & Zou, X. (2010). Advances and Challenges in Protein-Ligand Docking. *International Journal of Molecular Sciences*, *11*(8), 3016–3034. doi:10.3390/ijms11083016 PMID:21152288

Ian, I.-F., Pablo, C. V. J., Normande, C.-I., & Absalom, Z.-C. et al. (2013). Identification of pharmacological targets combining docking and molecular dynamics simulations. *American Journal of Agricultural and Biological Sciences*, *8*(1), 89–106. doi:10.3844/ajabssp.2013.89.106

Katritch, V., Jaakola, V.-P., Lane, J. R., Lin, J., IJzerman, A. P., Yeager, M., & Abagyan, R. et al. (2010). Structure-Based Discovery of Novel Chemotypes for Adenosine A_{2A}Receptor Antagonists. *Journal of Medicinal Chemistry*, *53*(4), 1799–1809. doi:10.1021/jm901647p PMID:20095623

Kitchen, D. B., Decornez, H., Furr, J. R., & Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nature Reviews. Drug Discovery*, *3*(11), 935–949. doi:10.1038/nrd1549 PMID:15520816

Korb, O., Stutzle, T., & Exner, T. E. (2006). PLANTS: Application of Ant Colony optimization to Structure-Based Drug Design. *ANTS 2006*. *LNCS*, *4150*, 247–258.

Kroeger, M. B., Rizzo, R. C., Hose, B. M., Tirado-Rives, J., & Smith, R. H. (2003). Molecular modeling calculations of HIV-1 reverse transcriptase nonnucleoside inhibitors: Correlation of binding energy with biological activity for novel 2-aryl-substituted benzimidazole analogues. *Journal of Medicinal Chemistry*, *46*(10), 1940–1947. doi:10.1021/jm020271f PMID:12723956

Krovat, E. M., Steindl, T., & Langer, T. (2005). Recent advances in docking and scoring. *Current Computer-aided Drug Design*, *1*(1), 93–102. doi:10.2174/1573409052952314

Kumalo, H. M., Bhakat, S., & Soliman, M. E. S. (2015). Theory and applications of covalent docking in drug discovery: Merits and pitfalls. *Molecules (Basel, Switzerland)*, *20*(2), 1984–2000. doi:10.3390/molecules20021984 PMID:25633330

Kumar, A., Goyal, R., Kumar, S., Jain, S., Jain, N., & Kumar, P. (2015). Estrogenic and Anti-Alzheimer's studies of *Zingiber officinalis* as well as *Amomum subulatumRoxb*.: The success story of dry techniques. *Medicinal Chemistry Research*, *24*(3), 1089–1097. doi:10.1007/s00044-014-1166-y

Kumar, A., Kumar, S., Jain, S., Kumar, P., & Goyal, R. (2013). Study of binding of pyridoacridine alkaloids on Topoisomerase II using *In-Silico* tools. *Medicinal Chemistry Research*, *22*(11), 5431–5441. doi:10.1007/s00044-013-0496-5

Lafleur, K., Huang, D., Zhou, T., Caflisch, A., & Nevado, C. (2009). Structure based optimization of potent and selective inhibitors of the tyrosine kinase erythropoietin producing human hepatocellular carcinoma receptor B4 (EphB4). *Journal of Medicinal Chemistry*, *52*(20), 6433–6446. doi:10.1021/jm9009444 PMID:19788238

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Laurini, E., Dal Col, V., Mamolo, M. G., Zampieri, D., & Pricl, S. et al. (2011). Homology model and docking-based virtual screening for ligands of the sigma (1) receptor. *ACS Medicinal Chemistry Letters*, *2*(11), 834–839. doi:10.1021/ml2001505 PMID:24900272

Leach, A. R., Shoichet, B. K., & Peishoff, C. E. (2006). Prediction of Protein-Ligand Interactions. Docking and Scoring: Successes and Gaps. *Journal of Medicinal Chemistry*, *49*(20), 5851–5855. doi:10.1021/jm060999m PMID:17004700

Legge, F. S., Budi, A., Treutlein, H., & Yarovsky, I. (2006). Protein flexibility: Multiple molecular dynamics simulations of insulin chain B. *Biophysical Chemistry*, *119*(2), 146–157. doi:10.1016/j.bpc.2005.08.002 PMID:16129550

Lyne, P. D., Lamb, M. L., & Saeh, J. C. (2006). Accurate prediction of the relative potencies of members of a series of kinase inhibitors using molecular docking and MM-GBSA scoring. *Journal of Medicinal Chemistry*, *49*(16), 4805–4808. doi:10.1021/jm060522a PMID:16884290

Ma, B., Shatsky, M., Wolfson, H. J., & Nussinov, R. (2002). Multiple diverse ligands binding at a single protein site: A matter of pre-existing populations. *Protein Science*, *11*(2), 184–197. doi:10.1110/ps.21302 PMID:11790828

Mahajan, A., & Gill, N. S. (2014). A review on molecular docking. *International Journal of Recent Adv Pharm Res.*, *4*(2), 1–7.

Mahasenan, K. V., & Li, C. (2012). Novel inhibitor discovery through virtual screening against multiple protein conformations generated via ligand-directed modeling: A maternal embryonic leucine zipper kinase example. *Journal of Chemical Information and Modeling*, *52*(5), 1345–1355. doi:10.1021/ci300040c PMID:22540736

McConkey, B. J., Soboley, V., & Edelmen, M. (2002). The performance of current methods in ligand–protein docking. *Current Science*, *83*(7), 845–956.

Mohan, V., Gibbs, A. C., Cummings, M. D., Jaeger, E. P., & DesJarlais, R. L. (2005). Docking: Successes and Challenges. *Current Pharmaceutical Design*, *11*(3), 323–333. doi:10.2174/1381612053382106 PMID:15723628

Motto, I., Bordogna, A., Soshilov, A. A., Denison, M. S., & Bonati, L. (2011). New aryl hydrocarbon receptor homology model targeted to improvedocking reliability. *Journal of Chemical Information and Modeling*, *51*(11), 2868–2881. doi:10.1021/ci2001617 PMID:21981577

Nandigam, R. K., Kim, S., Singh, J., & Chuaqui, C. (2009). Position specific interaction dependent scoring technique for virtual screening basedon weighted protein–ligand interaction fingerprint profiles. *Journal of Chemical Information and Modeling*, *49*(5), 1185–1192. doi:10.1021/ci800466n PMID:19415918

Nguyen, E. D., Norn, C., Frimurer, T. M., & Meiler, J. (2013). Assessment and Challenges of Ligand Docking into Comparative Models of G-Protein Coupled Receptors. *PLoS ONE*, *8*(7), e67302. PMID:23844000

Ogungbe, I. V., & Setzer, W. N. (2013). *In-silico* Leishmania target selectivity of antiparasitic terpenoids. *Molecules (Basel, Switzerland)*, *18*(7), 7761–7847. doi:10.3390/molecules18077761 PMID:23823876

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Ouyang, X., Zhou, S., Ge, Z., Li, R., & Kwoh, C. K. (2013). CovalentDock Cloud: A web server for automated covalent docking. *Nucleic Acids Research*, 41. PMID:23677616

Perez-Nueno, V. I., Rabal, O., Borrell, J. I., & Teixido, J. (2009). APIF: A newinteractionfingerprint based on atom pairs and its application tovirtual screening. *Journal of Chemical Information and Modeling*, *49*(5), 1245–1260. doi:10.1021/ci900043r PMID:19364101

Ranson, M., Mansoor, W., & Jayson, G. (2002). ZD1839 (IRESSA): A selective EGFR-TK inhibitor. *Expert Review of Anticancer Therapy*, *2*(2), 161–168. doi:10.1586/14737140.2.2.161 PMID:12113238

Rao, B. C., Subramanian, J., & Sharma, S. D. (2009). Managing protein flexibility in docking and its applications. *Drug Discovery Today*, *14*(7-8), 394–400. doi:10.1016/j.drudis.2009.01.003 PMID:19185058

Reddy, A. S., Pati, S. P., Kumar, P. P., Pradeep, H. N., & Sastry, G. N. (2007). Virtual Screening in Drug Discovery– A Computational Perspective. *Current Protein & Peptide Science*, *8*, 329–351. doi:10.2174/138920307781369427 PMID:17696867

Roy, K., Kar, S., & Das, R. N. (2015). *Understanding the basics of qsar for applications in pharmaceutical sciences and risk assessment*. Elsevier Inc.

Sali, A., & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*, *234*(3), 779–815. doi:10.1006/jmbi.1993.1626 PMID:8254673

Schapira, M., Raaka, B. M., Samuels, H. H., & Abagyan, R. (2001). *In-silico* discovery of novel Retinoic Acid Receptor agonist structures. *BMC Structural Biology*, *1*(1), 1. doi:10.1186/1472-6807-1-1 PMID:11405897

Schulz-Gasch, T. & Stahl, M. (2004). Scoring functions for protein–ligand interactions: a critical perspective. *Drug discovery today: technologies | lead optimization, 1*(3), 231-239.

Seifert, M. H. (2009). Robust optimization of scoring functions for a target class. *Journal of Computer-Aided Molecular Design*, *23*(9), 633–644. doi:10.1007/s10822-009-9276-1 PMID:19471858

Singh, J., Petter, R. C., Baillie, T. A., & Whitty, A. (2011). The resurgence of covalent drugs. *Nature Reviews. Drug Discovery*, *10*(4), 307–317. doi:10.1038/nrd3410 PMID:21455239

Smith, J. (2005). Erlotinib: Small-molecule targeted therapy in the treatment of non-small-cell lung cancer. *Clinical Therapeutics*, *27*(10), 1513–1534. doi:10.1016/j.clinthera.2005.10.014 PMID:16330289

Sousa, S. F., Ribeiro, A. J. M., Coimbra, J. T. S., Neves, R. P. P., Martins, S. A., Moorthy, N. S. H. N., & Ramos, M. J. et al. (2013). Protein-ligand docking in the new millennium– a retrospective of 10years in the field. *Current Medicinal Chemistry*, *20*(18), 2296–2314. doi:10.2174/0929867311320180002 PMID:23531220

Stark, J. L., & Powers, R. (2012). Application of NMR and Molecular Docking in Structure-Based Drug Discovery. *Topics in Current Chemistry*, *326*, 1–34. doi:10.1007/128_2011_213 PMID:21915777

Stuart, B. G., Coxon, J. M., Morton, J. D., Abell, A. D., McDonald, D. Q., Aitken, S. G., & Bickerstaffe, R. et al. (2011). Molecular modeling: A search for a calpain inhibitor as a new treatment for cataractogenesis. *Journal of Medicinal Chemistry*, *54*(21), 7503–7522. doi:10.1021/jm200471r PMID:21955158

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Tahir, R. A., Sehgal, S. A., Khattak, N. A., Khan Khattak, J. Z., & Mir, A. (2013). Tumor necrosis factor receptor superfamily 10B (TNFRSF10B): An insight from structure modeling to virtual screening for designing drug against head and neck cancer. *Theoretical Biology & Medical Modelling*, *10*(1), 38. doi:10.1186/1742-4682-10-38 PMID:23724937

Tautermann, C. S. (2014). GPCR structures in drug design, emerging opportunities with new structures. *Bioorganic & Medicinal Chemistry Letters*, *24*(17), 4073–4079. doi:10.1016/j.bmcl.2014.07.009 PMID:25086683

Taylor, R. D., Jewsbury, P. J., & Essex, J. W. (2002). A review of protein-small molecule docking methods. *Journal of Computer-Aided Molecular Design*, *16*(3), 151–166. doi:10.1023/A:1020155510718 PMID:12363215

Thomsen, R. (2003). Flexible ligand docking using evolutionary algorithms: Investigating the effects of variation operators and local search hybrids. *Bio Systems*, *72*(1-2), 57–73. doi:10.1016/S0303-2647(03)00135-7 PMID:14642659

Totrov, M., & Abagyan, R. (2008). Flexible ligand docking to multiple receptor conformations: A practical alternative. *Current Opinion in Structural Biology*, *18*(2), 178–184. doi:10.1016/j.sbi.2008.01.004 PMID:18302984

Tripathi, S. K., Muttineni, R., & Singh, S. K. (2013). Extra precision docking, free energy calculation and molecular dynamics simulation studies of CDK2 inhibitors. *Journal of Theoretical Biology*, *334*, 87–100. doi:10.1016/j.jtbi.2013.05.014 PMID:23727278

Tuccinardi, T., Botta, M., Giordano, A., & Martinelli, A. (2010). Protein kinases: Docking and homology modeling reliability. *Journal of Chemical Information and Modeling*, *50*(8), 1432–1441. doi:10.1021/ci100161z PMID:20726600

Vass, M., Schmidt, E., Horti, F., & Keseru, G. M. (2014). Virtual fragment screening on GPCRS: A case study on dopamine D3 and histamine H4 receptors. *European Journal of Medicinal Chemistry*, *77*, 38–46. doi:10.1016/j.ejmech.2014.02.034 PMID:24607587

Vyas, V., Jain, A., Jain, A., & Gupta, A. (2008). Virtual screening: A fast tool for drug design. *Scientia Pharmaceutica*, *76*(3), 333–360. doi:10.3797/scipharm.0803-03

Wada, M., Kanamori, E., Nakamura, H., & Fukunishi, Y. (2011). Selection of *in-silico* drug screening results for G-protein-coupled receptors by using universal active probes. *Journal of Chemical Information and Modeling*, *51*(9), 2398–2407. doi:10.1021/ci200236x PMID:21848279

Wlodawer, A. (2002). Rational approaches to AIDS drug design through structural biology. *Annual Review of Medicine*, *53*(1), 595–614. doi:10.1146/annurev.med.53.052901.131947 PMID:11818491

Wong, C. F. (2008). Flexible ligand–flexible protein docking in protein kinase systems. *Biochimica et Biophysica Acta*, *1784*(1), 244–251. doi:10.1016/j.bbapap.2007.10.005 PMID:17996742

Yuriev, E., Agostino, M., & Ramsland, P. A. (2011). Challenges and advances in computational docking: 2009 in review. *Journal of Molecular Recognition*, *24*(2), 149–164. doi:10.1002/jmr.1077 PMID:21360606

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Yuriev, E., & Ramsland, P. A. (2013). Latest developments in molecular docking: 2010–2011 in review. *Journal of Molecular Recognition*, *26*(5), 215–239. doi:10.1002/jmr.2266 PMID:23526775

Zhang, X., Wang, H., Li, Y., Cao, R., Zhong, W., Zheng, Z., & Li, S. et al. (2013). Novel substituted heteroaromatic piperazine and piperidine derivatives as inhibitors of human enterovirus 71 and coxsackievirus a16. *Molecules (Basel, Switzerland)*, *18*(5), 5059–5071. doi:10.3390/molecules18055059 PMID:23629759

Zhou, T., Huang, D., & Caflisch, A. (2010). Quantum Mechanical Methods for Drug Design. *Current Topics in Medicinal Chemistry*, *10*(1), 33–45. doi:10.2174/156802610790232242 PMID:19929831

This work was previously published in Methods and Algorithms for Molecular Docking-Based Drug Design and Discovery edited by Siavoush Dastmalchi, Maryam Hamzeh-Mivehroud, and Babak Sokouti, pages 295-319, copyright year 2016 by Medical Information Science Reference (an imprint of IGI Global).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 32
Current Trends in Docking Methodologies

Shubhandra Tripathi
CSIR-Central Institute of Medicinal and Aromatic Plants, India

Amandeep Kaur Kahlon
International Centre for Genetic Engineering and Biotechnology, India

Akhil Kumar
CSIR-Central Institute of Medicinal and Aromatic Plants, India

Ashok Sharma
CSIR-Central Institute of Medicinal and Aromatic Plants, India

ABSTRACT

Molecular docking was earlier considered to predict the binding affinity of the receptor and ligand molecules. With the progress in computational power and developing approaches, new horizons are now opening for accurate prediction of molecular binding affinity. In the current book chapter, recent strategies for Computer-Aided Drug Designing (CADD) including virtual screening and molecular docking, encompassing molecular dynamics simulations, and binding free energy calculation methods are discussed. Brief overview of different binding free energy methods MMPBSA, MMGBSA, LIE and TI have also been given along with the recent Relaxed Complex Scheme protocol.

1. INTRODUCTION

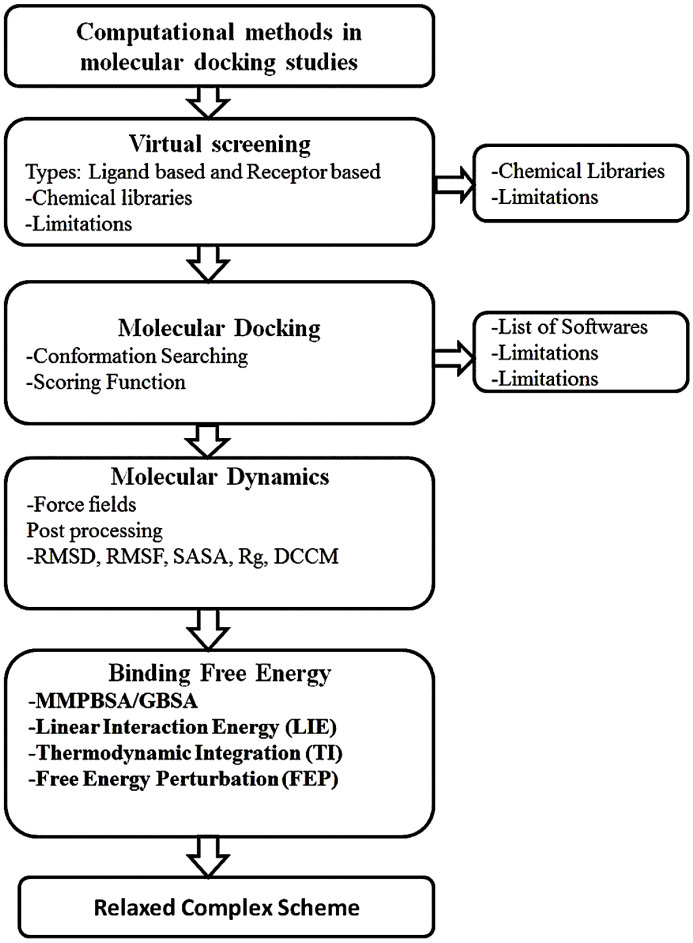
The potential of computer aided drug design was recognized earlier. Fortune magazine in 1981 published a cover article titled “Next Industrial Revolution: Designing Drugs by Computer at Merck” (Van Drie, 2007). Initially, high throughput screening (HTS) was used to find novel therapeutic molecules in drug discovery. It is the process of testing a large number of diverse chemical structures against disease targets to identify ‘hits’(Liu, Li, & Hu, 2004). Later, *in silico* techniques have acquired the competitive edge to filter out the lead molecules. Virtual screening is used for filtering molecules of interest from large chemical libraries. Screened molecules obtained as results are further evaluated using molecular docking. Moreover, the binding free energy methods viz. MMPBSA/MMGBSA, LIE, TI and FEP are used for stringent filtering from the molecular docking results. Lead molecules obtained from the *in silico* screening are used for biological evaluations (Figure 1).

DOI: 10.4018/978-1-5225-0549-5.ch032

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 1. Flowchart showing various sections in the book chapter



2. HIGH THROUGHPUT SCREENING (HTS)

HTS is characterized by taking the ligand-target interactions as the principle, leading to a higher information harvest (Liu et al., 2004). HTS lists contains a large percentage of false positives, making follow-up assays necessary to distinguish active from inactive substances (Jenkins, Kao, & Shapiro, 2003). HTS hit rate can be increased by using advance computational methods. Computational methods have several advantages over HTS as they are less time consuming, cost effective and require minimal compound design or prior knowledge. For example, researchers have used computational methods along with HTS to screen inhibitors of tyrosine phosphatase-1B, an enzyme implicated in diabetes. Virtual screening yielded nearly 35% hit rate as compared to 0.021% hit rate produced with HTS, demonstrating the power of computer aided drug designing (CADD) in drug discovery (Doman et al., 2002). Traditionally, HTS produces poor hit rate but in combination with CADD, its efficiency is greatly enhanced due to removal of false positives in less time and at lower computational cost.

One of the most remarkable uses of CADD in drug discovery process is the discovery of new targets for already existing drugs, a vice-versa approach. A few examples of approved drugs that owe their discovery

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

to CADD includes: carbonic anhydrase inhibitor dorzolamide, approved in 1995 (Vijayakrishnan, 2009); the angiotensin-converting enzyme (ACE) inhibitor captopril, approved in 1981 as an antihypertensive drug (Talele, Khedkar, & Rigby, 2010).

2.1 Virtual Screening: Filter Efficient Compound from Large Chemical Space

The term "virtual screening" appeared for the first time in a peer-reviewed publication dating 1997. Virtual screening is the filtering of large chemical compound libraries and is extensively used in computer aided drug design. The main concept of virtual screening is to sort active molecules from a large chemical library which may bind to the target protein with reduced off target effects (Dror, Shulman-Peleg, Nussinov, &Wolfson, 2004). Nowadays, CADD is playing a crucial role in research and development of novel drugs. Different steps includes: virtual screening of large database, binding affinity estimation between protein ligand through molecular docking, activity predication of ligands, dynamical interaction of ligand with the protein via molecular dynamics and binding free energy calculations. ADMET is also an important part of CADD analysis. These components may be combined into three major steps in drug discovery:

1. Screening of large chemical libraries into smaller sets of predicted active compounds that can be tested in real time i.e. virtual screening and molecular docking studies.
2. Optimization of lead compounds, to increase its affinity using in silico methods viz. molecular dynamics simulations and binding free energy estimations and drug metabolism and pharmacokinetics (DMPK) properties estimation including absorption, distribution, metabolism, excretion and toxicity potential (ADMET).
3. Design novel compounds, either by "growing" starting molecule's functional group at a time or by combining together fragments into novel chemotypes.

Due to limitation of performing biological assay for vast number of commercially available compounds, compound selection based on some reasonable criteria is required. The process must be fast enough to screen large databases containing millions of compounds along with maintaining accuracy. Virtual screening fulfills that requirement and may also be called as fast docking process. The basic difference is that docking calculations used for drug designing are for maximum accuracy as compared to docking calculations in virtual screening having fast calculations with limited accuracy. In virtual screening docking algorithms is designed to be extremely fast at the cost of accuracy because a large quantity of data is being searched. It is necessary to use techniques taking very small time in analyzing each molecule.

Similarity search is a tool used in virtual screening to find out structurally similar compounds to the known inhibitors against the target receptor. Similarity search tends to find only those compounds which are structurally very similar to the known active compounds. They don't search compounds with different backbones but focus only on those compounds having same or very similar backbone and functional groups. A substructure search can also be used to find compounds with an identical structural motif.

Virtual screening can be classified as ligand based virtual screening (LBVS) and structure based virtual screening (SBVS)(Dror et al., 2004). LBVS basically uses the information from the experimentally known active set of molecules (Jahn, Hinselmann, Fechner, & Zell, 2009) and SBVS calculates the interaction between the target receptor protein and ligands. LBVS can be performed by quantitative structure activity relationships (QSAR), similarity and substructure searching and 3D QSAR pharma-

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

cophore (Villoutreix et al., 2007). SBVS uses the scoring and energy function to rank the docked pose into the active site to identify best hits.

First step of VS is to identify the biological therapeutic drug targets. The molecular structure can be obtained from the protein database (Berman, 2000) or can be generated using computational methods such as homology modeling. An unexplored aspect of SBVS is the flexibility of target protein structure (Evers, Hessler, Matter, & Klabunde, 2005).To deal with flexibility, all the conformational structures of target protein are clustered and all molecules are screened against representative structures of each cluster. Soft docking approach is another method to deal with target flexibility. Soft docking implements the softening of van der Waals potentials allowing small overlaps between the ligand and receptor without large steric penalties (Jiang & Kim, 1991). However, this may increase the rate of false positives because more diverse structures are allowed to bind (Lavecchia & Di Giovanni, 2013). One of the recent methods focusing on receptor flexibility is Relaxed Complex Scheme (RCS). In this approach, receptor ensemble is generated from structures extracted from molecular dynamics (MD) simulations of receptor molecule and these ensemble structures are screened against ligand libraries (Amaro, Baron, & McCammon, 2008). The next step is to screen chemical libraries in databases against particular drug target.

2.2 Chemical Libraries

Chemical libraries used for screening are broadly classified into 3 classes:

1. General libraries that can be used to screen against any target.
2. Focused libraries that are designed for a family of related targets.
3. Targeted libraries that are specifically designed for a single target.

Several target-specific libraries are available in Cambridge Structure Database viz., kinase biased, GPCR-biased and ion channel-biased sets. To screen these libraries, molecules as the ligand are to be prepared in 2D string format so that we can convert it into 3D structure. Also, ligand must be checked for incorrect bond length/distance, angle and order. Ligand structures requires preparation via 3D co-ordinate generation, geometry optimization, conformational sampling, assigning proper isometric and protonation state, and minimized molecular flexibility (Cavasotto & Phatak, 2011;Anderson, 2012). Chemical compounds can also be narrowed down by applying drug like or rule of five i.e. Lipinski filter but it may remove important scaffolds having higher binding affinity. Some of the important public databases are listed in Table1.

Besides, some other natural molecules database like Interbioscreen database, Duke phytochemical database, and French national chemical library are also popular having chemical libraries.

Once the chemical library is prepared, in SBVS we need a computational method to screen the top hit from the database. To achieve this goal several algorithms and scoring function are used. Mathematical algorithms are used to evaluate the binding of the ligand into the active site of target/ receptor. Scoring function and other criteria are used for ranking ligands. Scoring functions are very important in the SBVS and can be classified as force filed, knowledge based and empirical (Krovat, Steindl, & Langer, 2005; Jain, 2006). Scoring functions are described in detail in molecular docking section. Best option is to use combination of scoring function but it demands higher time and computational cost. In VS one has to screen millions of compounds, so docking and scoring function are selected with optimal accuracy and fast screening.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 1. List of databases having chemical libraries

Database	Size	Type
ZINC	35 million	Annotated commercially available compounds
ChemDB	5 million	Annotated commercially available molecules
ChemSpider	34 million	Chemical structure database
Pubchem	30 million	Biologic activities of small molecules
DrugBank	4,800 drugs; 2,500 targets	Detailed drug data with comprehensive drug target Information
ChemBank	1,2 million	Biomedically relevant data and tools. Small molecules and small-molecule screens, and resources for studying the data
NCI Open Database	265,000	Compound libraries
ChEMBL	1 million	Chemicals extracted from literature, including target and bioactivity information
CoCoCo	7 million	Molecular structural information of commercial compounds from various chemical vendors by providing it in a ready-to-use format.

LBVS method uses set of active ligand and extract information from ligands necessary to bind. Ligands align itself to compute their shared pharmacophore. A pharmacophore is defined as the 3D arrangement of features that is crucial for a ligand molecule in order to interact with a target receptor in a specific binding site. Molecular docking interaction study is performed for lead molecules obtained from virtual screening. Utilizing molecular docking lead molecules are further filtered and those showing good binding affinity are processed further.

2.3 Limitations

Virtual screening plays a vital role in drug discovery. But it suffers from various limitations. The major limitations of the virtual screening and molecular docking are mentioned below.

1. The main and primarily goal of VS is to identify as many as possible diverse hits and not the most potent molecules. It may therefore, predict wrong binding pose due to biased algorithm and scoring function.
2. Currently, the role of solvent is not considered in virtual screening, while the water molecule have prominent role in hydrogen bond interaction.
3. Criteria for selecting the compounds after VS is ambiguous. Further, positive controls of negative data sets are rarely used in virtual screening.
4. Fundamental error may occur in converting molecular structure to another format. Various softwares have different protocols which are non-standardized. Serious concerns are regarding alternation in atomic coordinates, chirality, hybridization and partial charges.
5. One of the major challenges is generating an optimal set of conformations. Good coverage of conformational space is also important. Similarly, we have to remove high energy conformations, if they are not removed they will be present in the positive hits.
6. Flexibility of ligand is considered in VS but target flexibility is probably the most unexplored aspect of VS. It should be considered while doing SBVS.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

VS require careful preparation of the database, judicious parameter choices, and sensible compromises between the different goals one attempts to obtain. A series of government-funded initiatives which mandate deposition of screening data in public repositories and promote the use of open source software has resulted into availability of vast datasets including more bioactivity classes. This addresses many of the biases inherent in earlier benchmark collections. While VS will always remain a probabilistic game, the odds of success can be greatly enhanced through careful planning and attention to details.

3. MOLECULAR DOCKING

Molecular docking is the interaction study of protein and ligand, the interaction is defined in terms of binding affinity. Kuntz et al., 1982 developed first rigid body docking software, based on geometric complementarity of interacting molecules. Docking has now improved to greater extent having efficient scoring functions and accurate predictions.

Molecules in the nature interact with each other to form various interactions. This is a vital process in nature to form various molecular associations. Using in silico molecular docking methods, various factors can be analyzed for molecular interactions. The fundamental approach used in molecular docking comprises, conformational space search of the three dimensional space of the protein with the ligand molecule (docking) and predicting the affinity of these poses using scoring function (scoring)(Kitchen, Decornez, Furr, & Bajorath, 2004).

3.1 Conformational Searching

Several degrees of freedom are computed for the ligand molecule in molecular docking. There are total six translational and rotational degrees of freedom that are calculated for the conformations of ligand molecule relative to receptor. Various methods includes genetic algorithm based on evolutionary approach (e.g. Autodock, GOLD), incremental construction method as in FlexX or methods based on molecular mechanics force field (e.g. Glide). A ligand molecule docked in the protein active site is shown in figure 2 (Tripathi, Kumar, Kumar, Negi, & Sharma, 2015).

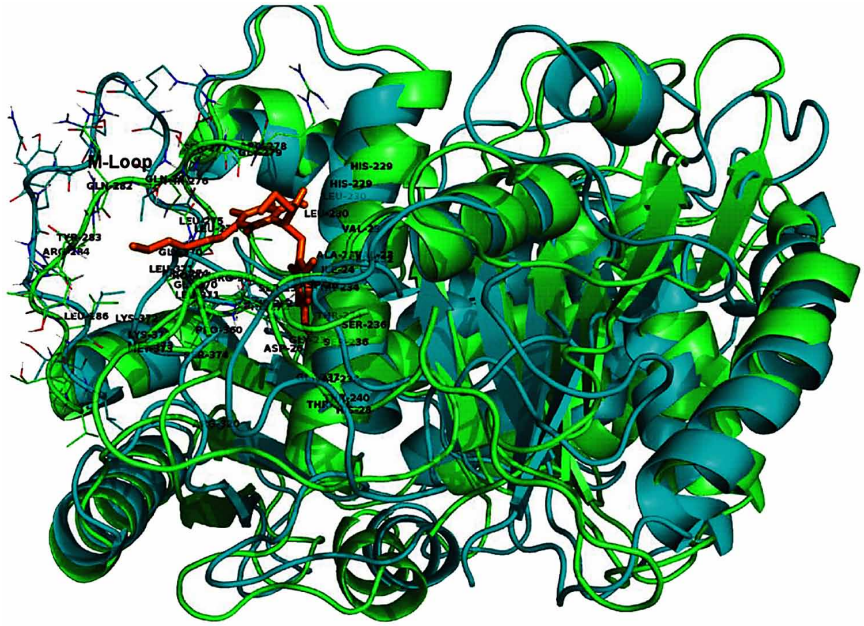
3.2 Scoring Function

Scoring functions are the mathematical value used to calculate the binding of the ligand molecule with receptor protein. Scoring functions are majorly of three types: force field scoring function, empirical scoring function and knowledge based scoring functions (Bortolato, Fanton, Mason, & Moro, 2013). Molecular mechanics force field based scoring function utilizes various energy terms to calculate the binding free energy. The widely used force fields are AMBER(Wang, Wolf, Caldwell, Kollman, & Case, 2004) and OPLS (Jorgensen &Tirado-Rives, 1988). Empirical scoring function uses regression method to predict the binding free energy. Coefficients are calculated from ligands with known experimental binding free energy. Similarly, in knowledge based scoring function, linear regression analysis or other methods like machine learning techniques are used to predict the binding free energy. Ligand conformations are ranked according to scoring function, and the best conformations are used for further analysis.

In real conditions, protein adopts different conformations upon ligand binding as compared to their unbound form. These changes in receptor conformations have significant impact in calculating the bind-

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 2. Protein-ligand docked complex image; ligand bound to the protein active site



ing free energy. Therefore, the ligand binding predictions using methods like docking and scoring are not very accurate. There are two widely used methods to overcome this limitation:

1. **Conformational Selection:** In this method all the conformations of the unbound receptor available are utilized in docking and scoring (Monod, Wyman, & Changeux, 1965;Ma, Kumar, Tsai, & Nussinov, 1999). Receptor ensemble is generated using multiple X-ray, NMR structure of the receptor protein or through multiple snapshots from MD simulations (Lin, Perryman, Schames, & McCammon, 2002).
2. **Induced Fit:** The configuration of receptor is shifted to a different configuration upon ligand binding interaction. The study of receptor flexibility is performed using various techniques including Molecular Dynamics (MD), Monte Carlo, Umbrella Sampling, Metadynamics, and accelerated MD.

Molecular docking is broadly classified into three categories on the basis of receptor flexibility while performing docking experiment-

1. Rigid docking, where docking of a rigid receptor molecule is performed with rigid ligand molecule.
2. Flexible docking, where receptor protein molecule behaves as a rigid entity while the ligand molecule has complete flexibility, its all rotational bonds can change internal torsions.
3. Full flexible docking, where the ligands have torsion angle flexibility while the selected amino acid residues of the protein active side chains also remain flexible.

Various docking softwares are available with different algorithm and scoring functions (Lavecchia & Di Giovanni, 2013). Some freely available softwares are listed in table2.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 2. List of molecular docking softwares

Software	Free for Academia	Details	Scoring Function
Autodock	Yes	Uses an interaction grid to account for receptor conformations and simulated annealing to account for ligand conformations. Available free of charge	Force field
DOCK	Yes	Docks small molecules, includes solvent effects, uses incremental construction for ligands. Available to academic users without charge	Force field or contact score
AutodockVina	Yes	AutoDockVina significantly improves the average accuracy of the binding mode predictions compared to AutoDock.	Force field
Molegro	Yes(30 days)	MolDock is based on a new heuristic search algorithm that combines differential evolution with a cavity prediction algorithm. The docking scoring function of MolDock is an extension of the piecewise linear potential (PLP) including new hydrogen bonding and electrostatic terms	Re-ranking scoring function
ZDOCK	Yes (Server)	Pair wise shape complementarity with desolvation and electrostatics	Pair wise shape complementarity base scoring function

For the structure-based drug designing 3D structure of the target protein is required. 3D structure can be derived by X-ray, NMR or one can model from the primary sequence. X-ray high-resolution crystallography data of target protein is more preferred than NMR data in the structural based drug design. In virtual screening structure of both protein and ligand are used to calculate interaction energies for all compounds that are tested. Ligand based CADD is generally preferred when no or little structural information is available, often for membrane protein targets (Sliwoski, Kothiwale, Meiler, & Lowe, 2013).

The docking software provides a ranked list of ligands. Using computer graphics, these are visually evaluated in complex with the target to assess goodness of fit, formation of key hydrogen bonds or hydrophobic interactions, surface complementarity and stability of the bound conformation of the ligand compared to the free conformation. Several top-scoring ligands are then purchased or synthesized and evaluated in the laboratory using *in vitro* assays with the target. Compounds that succeed as hits in the laboratory assays are then subjected to further rounds of *in silico* screening using more focused libraries of potential analogs. The structure of the complex comprising the target and hit molecules is often experimentally determined in order to validate the binding mode. To further investigate the binding affinity and stability of the receptor-ligand interaction molecular dynamics simulations and binding free energy calculations are performed. Molecular dynamics investigates the conformational stability of the complex and binding free energy estimates stringent calculation of binding affinity.

3.3 Limitations

To predict accurate binding affinities, it will be necessary to go beyond predicting a single “dominant” conformation of the ligand docked with the protein. Major limitations of molecular docking are-

1. Binding free energy is not driven by a single conformation, but rather by the free energy landscape.
2. It is the shape of the energy landscape that is crucial, the shape and width of the minima influences entropies.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

3. Entropies are key contributors to binding thermodynamics and are not observable in single bound structures.
4. Multiple ligand poses, protein conformations, strain energies, changes in water structure, and solvation and protonation all play roles in binding affinity. And none of these are observable in single structures.
5. Computational tools can help provide insight into the unseen landscape, so those doing crystallographic studies may want to complement their work by using computational tools to explore this landscape. And those relying on crystallographic data (e.g., in a drug design context) should be aware that there are various binding possibilities that might not be captured in a single crystal structure.

4. MOLECULAR DYNAMICS (MD) SIMULATION

Molecular dynamics simulation provides an important tool for validating the molecular docking results. Molecular dynamics simulation is to imitate the real conditions including solvent molecules, charge neutralizing by counter ions addition and then production MD to generate MD trajectory, resembling real time conformations of the system. Molecular dynamics imbibes the flexibility of the receptor molecules as well as the ligand molecule and helps in probing the interaction between them. In contrast with flexible docking where the protein flexibility remains limited, in MD simulations complete flexibility is maintained for the system. In MD of protein-ligand docked complex, ligand binding may cause subtle changes in the side chains rotation or may lead to large deviation in the backbone structure of the protein (Sinko, Lindert, & McCammon, 2013).Various conformational snapshots of different poses during the production run are saved in MD trajectories. All conformational changes in the protein and ligand can be assessed using the molecular dynamics trajectories. These MD trajectories are usually generated in nanosecond scale (usually 10-100 ns), containing multiple snapshots used for conformational investigations and binding free energy calculations. Molecular Dynamics simulations were first proposed based on classical law of motion (F= mxa) to describe atomic motions (McCammon, Gelin, & Karplus, 1977). Earlier complex quantum-mechanical calculations were utilized to define atomic motions.

In MD simulations all the forces acting on each atom of a biomolecular system are calculated using Newtonian physics. The energy contribution for every atom includes bonded interactions and non-bonded interactions(Cornell et al., 1995). The non-bonded interactions includes electrostatic interactions calculated using Columb's law, and vander Waals interactions using Lennard Jones Potentials (L-J potential)(Jones, 1924). The bonded interaction incorporates bond distances, bond angles, dihedral angles and improper angles, which are calculated using classical physical equations (Cornell et al., 1995). The chemical bonds and bond angles are represented as virtual springs, and dihedral angles are modeled as sinusoidal functions. All energy terms are optimized to fit the quantum-mechanical calculations to reproduce the real behavior of atomic motions of molecules (Durrant & McCammon, 2011).

4.1 Force Fields

All the optimized energy terms cumulatively corresponds to the "force field', describing the contribution of various atomics forces that govern molecular dynamics. Various widely used force fields for molecular dynamics simulation include Assisted Model Building with Energy Refinement (AMBER)

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

(Wang et al., 2004), GROningen MAchine for Chemical Simulations (GROMACS) (Berendsen, van der Spoel, & van Drunen, 1995), Chemistry at HARvard Molecular Mechanics (CHARMM) (Brooks et al., 1983; Vanommeslaeghe et al., 2009) and Optimized Potential for Liquid Simulations (OPLS) (Jorgensen & Tirado-Rives, 1988). These force fields have different parameters and are optimized for different biomolecules viz. protein, nucleic acids, and small chemical molecules.

All the forces acting on each of the atom of system are calculated and then the position of atoms are moved according to direction of net applied force using Newton's law of motion. The time step in which the relative positions of atom are determined is usually 2 femtosecond (10-15 second). Then the relative positions of atoms of biomolecule are estimated in further steps and the generated snapshots are saved. These snapshots in a long timescale lead to formation of MD trajectories. MD trajectories are further used for analyzing various parameters like RMSD, RMSF etc.

Most commonly used softwares for molecular dynamics simulation are AMBER (Case et al., 2005; Wang et al., 2004), CHARMM (Brooks et al., 1983), GROMACS (Van Der Spoel et al., 2005; Hess, Kutzner, van der Spoel, & Lindahl, 2008) and NAMD (Kalé et al., 1999; Phillips et al., 2005).

4.2 MD Analysis

From MD simulation conformation trajectory is generated containing multiple snapshots of the system. Generated MD trajectory is utilized for various analysis to study the stability of the system. The secondary structure component or any other specific fluctuation in the 3-dimensional structure are analysed using various statistical methods. Methods used for analyzing the stability and structural changes are discussed below.

4.2.1 Root Mean Square Deviation (RMSD)

Root Mean Square Deviation (RMSD) of the atoms in a molecule is calculated with respect to a reference structure. The atoms in the molecule are fitted to the reference structure by roto-translational least square fitting (Karpen, Tobias, & Brooks, 1993). RMSD is calculated for the snapshot of protein molecule at two time point t1 and t2. Protein conformation at t1 is the reference structure and the RMSD of conformation at t2 is to be calculated by taking the average of the square of their atomic coordinate's difference. An example RMSD plot is shown in Figure 3 (Tripathi et al., 2015).

4.2.2 Root Mean Square Fluctuations (RMSF)

Root Mean Square Fluctuation (RMSF) is the standard deviation of the atomic positions in the MD trajectory fitting to a reference structure. Atomic fluctuations of individual amino acid in the protein are depicted by RMSF. RMSF calculations are used to investigate the higher flexibility shown in the loop regions of the receptor proteins. A RMSF plot is depicted in figure 4 (Tripathi et al., 2015).

4.2.3 Solvent Accessible Surface Area (SASA)

Solvent Accessible Surface Area (SASA) is calculated from the MD trajectories. SASA includes hydrophobic, hydrophilic and total solvent accessible surface area of the protein molecule. The calculated surface area is the canonical surface area (Eisenhaber, Lijnzaad, Argos, Sander, & Scharf, 1995). The

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 3. Plot showing RMSD graph, receptor protein (TUB), receptor protein with known ligand (TUB + Ctrl), receptor protein with other ligands (TUB + lig1 and lig2)

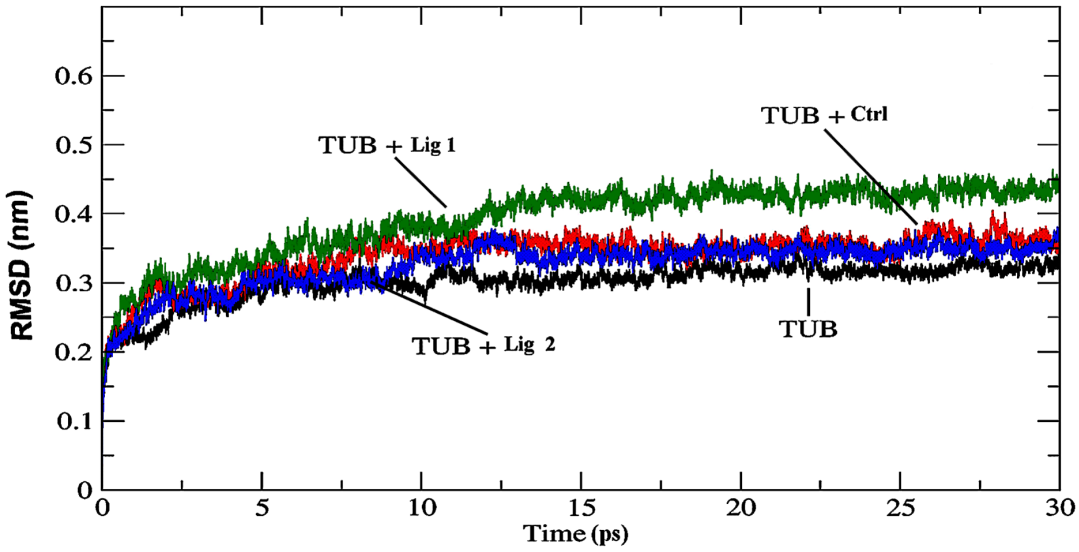
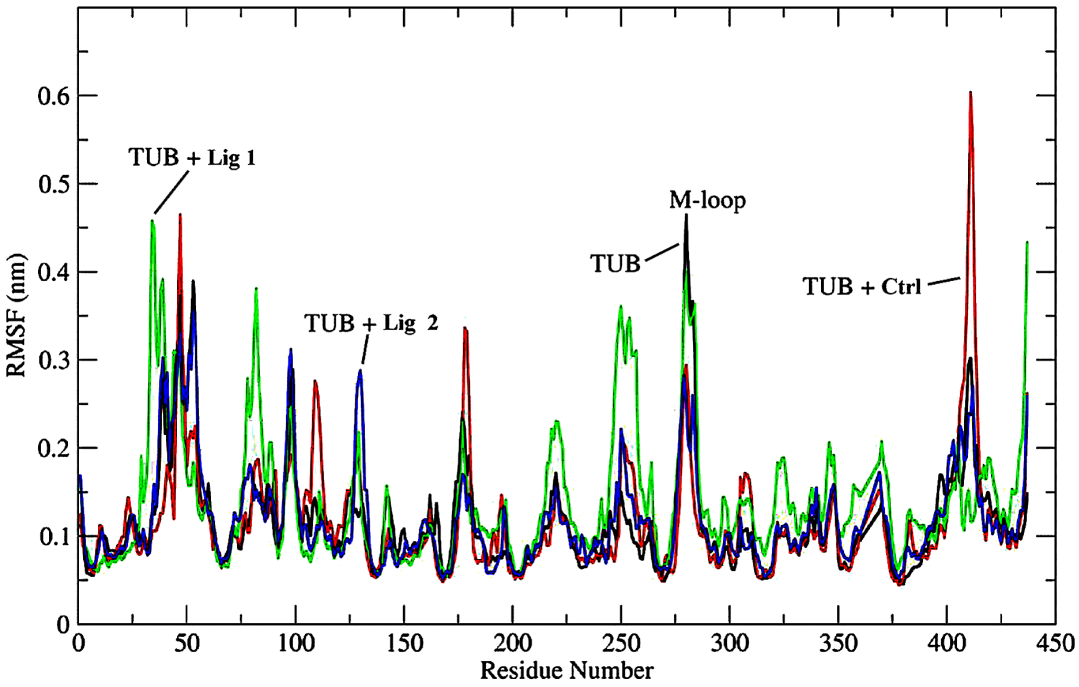


Figure 4. Plot showing RMSF graph, receptor protein unbound (TUB), receptor protein with known ligand (TUB + Ctrl), receptor protein with other ligands (TUB + lig1 and lig2)



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

extent to which the amino acids interact with the solvent and the protein core is proportional to the surface area exposed to the solvent (Durham, Dorr, Woetzel, Staritzbichler, & Meiler, 2009).

4.2.4 Radius of Gyration (Rg)

Radius of gyration (Rg) represents the compactness of the protein structure. The Rg of a protein is derived in two steps, deriving the center of mass (R_c) of the 3-D coordinates and then the Rg was calculated for the estimation of the compactness of protein. Rg helps to probe the folding of the protein, a higher value of Rg suggest loose packing of various conformations. Stable Rg value from MD trajectories presents that unique stable spatial structure of protein is retained without any structural unfolding of protein (Lobanov, Bogatyreva, & Galzitskaya, 2008).

4.2.5 Hydrogen Bonding

The number of hydrogen bond interactions between protein and ligand, and the time span of individual hydrogen bond i.e. H-bond occupancy can also be calculated from the MD trajectories.

4.2.6 Domain Cross Correlation Map (DCCP)

DCCP analysis is performed to identify the protein domains from the generated MD trajectories (Arnold & Ornstein, 1997). It has a prominent role for the identification of domain in the protein structure. The collective domain motions in the protein structure and cross correlation between domains are analyzed along with the involvement of the active site in such motions (Luo & Bruice, 2002).

4.2.7 Essential Dynamics

All the collective motions of protein can be classified into two subspaces essential component containing few motions comprising of the most atomic fluctuations and the remaining component having very limited motions i.e. “physically constrained”(Amadei, Linssen & Bexerendsen, 1993). These essential components of the protein motions from the MD trajectories are responsible for overall fluctuations in the protein. Normal mode analysis (NMA) and Principle component analysis (PCA) are the two invaluable tools for studying conformational changes in proteins (Skjaerven, Martinez, & Reuter, 2011).

5. BINDING FREE ENERGY

Compared with molecular docking, MD simulations of protein-ligand docked complex have both protein and ligand in flexible manner. Earlier in docking programs priority was given to speed rather than accuracy while scoring but more accurate techniques are now available for the estimation of the binding free energy from MD trajectories. These advanced techniques includes molecular mechanics/Poisson-Boltzmann and surface area (MM/PB-SA) (P. A. Kollman et al., 2000), thermodynamic integration (TI) (P. Kollman, 1993; Adcock & McCammon, 2006), free energy perturbation(FEP) (P. Kollman, 1993; Kim et al., 2006), and linear interaction energy (LIE) (Johan Aqvist, Luzhkov, & Brandsdal, 2002). These are the sophisticated techniques to calculate the binding affinity between the protein and the ligand. TI

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

and FEP are the most rigorous techniques but they are computationally too expensive. MMPBSA/GBSA methods are now days widely used for estimation of binding free energy with lower computational costs than TI and FEP. A brief account of each technique is given below.

5.1 MMPBSA and MMGBSA Methods

Molecular Mechanics Poisson Boltzman's Surface Area (MMPBSA) or Generalised Born Suface Area (MMGBSA) methods are utilized to obtain the relative free energies of binding, mostly used for comparing ligands binding affinity. This is the post-processing step after MD simulation, the representative snapshots from the stable trajectories are used to calculate the free energy change. It has three components, molecular mechanics energy calculations, Poisson Boltzman or Generalised Born method and estimation of surface area. The molecular mechanics component is used to calculate the energy difference in the gas phase. The solvation free energy component including both polar and non-polar solvation is calculated using the implicit solvent. Poisson Boltzman and Generalised Born are the two methods, used for calculating the polar solvation free energy component. The non-polar free energy component is calculated by estimating the surface area component. The entropy contribution to the total free energy was estimated further by the normal mode analysis or quasiharmonic approximation method. Ensemble based screening utilizes various receptor conformations for docking. Conformations can be determined from X-ray or NMR structures, Monte Carlo simulation, and molecular dynamics simulation or enhanced sampling methods.

5.2 Linear Interaction Energy

Linear interaction energy method estimates changes in the electrostatic and van der Waals interaction energy between the ligand and the surrounding environment, for the receptor bound ligand and the ligand in solution based on MD simulation (J. Aqvist, Medina, & Samuelsson, 1994). The binding free energy is estimated by calculating difference in electrostatic and van der Waals energies in a linear free energy equation. Difference in the solvent accessible surface area between bound and unbound ligand is also considered in estimating the free energy term (van Lipzig et al., 2004). These calculations are performed from the snapshots retrieved from the equilibrated MD trajectories.

5.3 Free Energy Perturbation (FEP) and Thermodynamics Integration (TI)

Free energy perturbation and Thermodynamics integration are rigorous and computationally expensive methods requiring higher computational cost. It's due to two main reasons, first the explicit solvent is used and second is to calculate the difference in free energy of two states on the basis of molecular dynamics simulations carried out at intermediate points along the transition path from one state to another state. Therefore, multiple molecular dynamics simulations are performed. The computational cost is reduced by using the implicit solvent approach to the end point states.

The advent of novel parallel computing architecture Compute Unified Device Architechture (CUDA) which utilizes Graphical Processing Units (GPUs), enhanced the sampling of the molecular modeling softwares. Future improvements and developments in the computational power will enhance the sampling of molecular docking, long time molecular dynamics simulations and stringent binding free estimation methods.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

6. RELAXED COMPLEX SCHEME

Lin et al., 2002 proposed a novel computational methodology, accommodating the idea of receptor flexibility, for drug designing called Relaxed Complex Scheme (RCS). Later Amaro et al., 2008 improved the RCS method by including virtual screening, increasing computational efficiency and more stringent binding calculations for computer aided drug designing.

RCS is a benchmarking computational protocol for drug designing, including docking algorithms with dynamical structural information obtained through molecular dynamics (MD) simulations, accounting for flexibility of ligand as well as receptor under study. The conformational flexibility for both receptor and ligand are taken into consideration in receptor ensemble and ligand ensemble then flexible docking studies were performed followed by molecular dynamics (MD) simulations. Binding free energy are calculated using implicit solvent methods viz. MMPBSA (Damm & Carlson, 2007; Homeyer & Gohlke, 2012), MMGBSA or coupling-parameter approaches viz. thermodynamics integration (TI) and free energy perturbation (FEP) including extensive sampling of receptor, ligand, and solvents.

For a novel protein target, virtual screening approaches are utilized to identify novel potential inhibitors. In previous RCS version receptor ensemble were retrieved from the snapshots at regular time intervals from MD trajectories, resulting in thousands of receptor structures, most of them are conformational redundant. While in advanced RCS clustering method based on pair-wise RMSD of the aligned structures was utilized. For receptors having known binding region the clustering algorithm focuses on the subset of residues involved in receptor binding site. The receptor ensemble can also be derived from collection of independent X-ray or NMR experimental structures (Damm & Carlson, 2007; Huang & Zou, 2007). In this way, 5-10 non-redundant conformations representing the receptor ensembles were generated, which were further used for docking studies and binding free energy estimates. Ligand ensemble will be taken from the various databases, also all the rotations in the ligand bonds will be permitted while performing the docking studies.

Molecular docking is performed for the receptor ensemble and the ligand database with Autodock open-source software (Goodsell, Morris, & Olson, 1996). In the Post-Processing step binding free energy was further evaluated using MMPBSA, single step Perturbation, Linear Interaction Energy (LIE), Free Energy Perturbation (FEP) and Thermodynamics Integration (TI) methods.

REFERENCES

Adcock, S. A., & McCammon, J. A. (2006). Molecular dynamics: Survey of methods for simulating the activity of proteins. *Chemical Reviews*, *106*(5), 1589–1615. doi:10.1021/cr040426m PMID:16683746

Amadei, A., Linssen, A. B., & Berendsen, H. J. (1993). Essential dynamics of proteins. *Proteins*, *17*(4), 412–425. doi:10.1002/prot.340170408 PMID:8108382

Amaro, R. E., Baron, R., & McCammon, J. A. (2008). An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *Journal of Computer-Aided Molecular Design*, *22*(9), 693–705. doi:10.1007/s10822-007-9159-2 PMID:18196463

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Anderson, A. C. (2012). Structure-Based Functional Design of Drugs: From Target to Lead Compound. In V. Espina & L. A. Liotta (Eds.), *Molecular Profiling* (Vol. 823, pp. 359–366). Totowa, NJ: Humana Press. doi:10.1007/978-1-60327-216-2_23

Aqvist, J., Luzhkov, V. B., & Brandsdal, B. O. (2002). Ligand binding affinities from MD simulations. *Accounts of Chemical Research*, *35*(6), 358–365. doi:10.1021/ar010014p PMID:12069620

Aqvist, J., Medina, C., & Samuelsson, J. E. (1994). A new method for predicting binding affinity in computer-aided drug design. *Protein Engineering*, *7*(3), 385–391. doi:10.1093/protein/7.3.385 PMID:8177887

Arnold, G. E., & Ornstein, R. L. (1997). Molecular dynamics study of time-correlated protein domain motions and molecular flexibility: Cytochrome P450BM-3. *Biophysical Journal*, *73*(3), 1147–1159. doi:10.1016/S0006-3495(97)78147-5 PMID:9284282

Berendsen, H. J. C., van der Spoel, D., & van Drunen, R. (1995). GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, *91*(1-3), 43–56. doi:10.1016/0010-4655(95)00042-E

Berman, H. M. (2000). The Protein Data Bank. *Nucleic Acids Research*, *28*(1), 235–242. doi:10.1093/nar/28.1.235 PMID:10592235

Bortolato, A., Fanton, M., Mason, J. S., & Moro, S. (2013). Molecular Docking Methodologies. In L. Monticelli & E. Salonen (Eds.), *Biomolecular Simulations* (Vol. 924, pp. 339–360). Totowa, NJ: Humana Press. doi:10.1007/978-1-62703-017-5_13

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, *4*(2), 187–217. doi:10.1002/jcc.540040211

Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., & Woods, R. J. (2005). The Amber biomolecular simulation programs. *Journal of Computational Chemistry*, *26*(16), 1668–1688. doi:10.1002/jcc.20290 PMID:16200636

Cavasotto, C. N., & Phatak, S. S. (2011). Docking methods for structure-based library design. *Methods in Molecular Biology (Clifton, N.J.)*, *685*, 155–174. doi:10.1007/978-1-60761-931-4_8 PMID:20981523

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., & Kollman, P. A. (1995). A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society*, *117*(19), 5179–5197. doi:10.1021/ja00124a002

Damm, K. L., & Carlson, H. A. (2007). Exploring experimental sources of multiple protein conformations in structure-based drug design. *Journal of the American Chemical Society*, *129*(26), 8225–8235. doi:10.1021/ja0709728 PMID:17555316

Doman, T. N., McGovern, S. L., Witherbee, B. J., Kasten, T. P., Kurumbail, R., Stallings, W. C., & Shoichet, B. K. (2002). Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. *Journal of Medicinal Chemistry*, *45*(11), 2213–2221. doi:10.1021/jm010548w PMID:12014959

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Dror, O., Shulman-Peleg, A., Nussinov, R., & Wolfson, H. J. (2004). Predicting molecular interactions in silico: I. A guide to pharmacophore identification and its applications to drug design. *Current Medicinal Chemistry*, *11*(1), 71–90. doi:10.2174/0929867043456287 PMID:14754427

Durham, E., Dorr, B., Woetzel, N., Staritzbichler, R., & Meiler, J. (2009). Solvent accessible surface area approximations for rapid and accurate protein structure prediction. *Journal of Molecular Modeling*, *15*(9), 1093–1108. doi:10.1007/s00894-009-0454-9 PMID:19234730

Durrant, J. D., & McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. *BMC Biology*, *9*(1), 71. doi:10.1186/1741-7007-9-71 PMID:22035460

Eisenhaber, F., Lijnzaad, P., Argos, P., Sander, C., & Scharf, M. (1995). The double cubic lattice method: Efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. *Journal of Computational Chemistry*, *16*(3), 273–284. doi:10.1002/jcc.540160303

Evers, A., Hessler, G., Matter, H., & Klabunde, T. (2005). Virtual screening of biogenic amine-binding G-protein coupled receptors: Comparative evaluation of protein- and ligand-based virtual screening protocols. *Journal of Medicinal Chemistry*, *48*(17), 5448–5465. doi:10.1021/jm050090o PMID:16107144

Goodsell, D. S., Morris, G. M., & Olson, A. J. (1996). Automated docking of flexible ligands: Applications of AutoDock. *Journal of Molecular Recognition: JMR*, *9*(1), 1–5. doi:10.1002/(SICI)1099-1352(199601)9:1<1::AID-JMR241>3.0.CO;2-6 PMID:8723313

Hess, B., Kutzner, C., van der Spoel, D., & Lindahl, E. (2008). GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation*, *4*(3), 435–447. doi:10.1021/ct700301q PMID:26620784

Homeyer, N., & Gohlke, H. (2012). Free Energy Calculations by the Molecular Mechanics Poisson−Boltzmann Surface Area Method. *Molecular Informatics*, *31*(2), 114–122. doi:10.1002/minf.201100135

Huang, S.-Y., & Zou, X. (2007). Ensemble docking of multiple protein structures: Considering protein structural variations in molecular docking. *Proteins*, *66*(2), 399–421. doi:10.1002/prot.21214 PMID:17096427

Jahn, A., Hinselmann, G., Fechner, N., & Zell, A. (2009). Optimal assignment methods for ligand-based virtual screening. *Journal of Cheminformatics*, *1*(1), 14. doi:10.1186/1758-2946-1-14 PMID:20150995

Jain, A. N. (2006). Scoring functions for protein-ligand docking. *Current Protein & Peptide Science*, *7*(5), 407–420. doi:10.2174/138920306778559395 PMID:17073693

Jenkins, J. L., Kao, R. Y. T., & Shapiro, R. (2003). Virtual screening to enrich hit lists from high-throughput screening: A case study on small-molecule inhibitors of angiogenin. *Proteins*, *50*(1), 81–93. doi:10.1002/prot.10270 PMID:12471601

Jiang, F., & Kim, S. H. (1991). "Soft docking": Matching of molecular surface cubes. *Journal of Molecular Biology*, *219*(1), 79–102. doi:10.1016/0022-2836(91)90859-5 PMID:2023263

Jones, J. E. (1924). On the Determination of Molecular Fields. II. From the Equation of State of a Gas. *Proceedings - Royal Society. Mathematical, Physical and Engineering Sciences*, *106*(738), 463–477. doi:10.1098/rspa.1924.0082

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Jorgensen, W. L., & Tirado-Rives, J. (1988). The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society*, *110*(6), 1657–1666. doi:10.1021/ja00214a001

Kalé, L., Skeel, R., Bhandarkar, M., Brunner, R., Gursoy, A., Krawetz, N., & Schulten, K. (1999). NAMD2: Greater Scalability for Parallel Molecular Dynamics. *Journal of Computational Physics*, *151*(1), 283–312. doi:10.1006/jcph.1999.6201

Karpen, M. E., Tobias, D. J., & Brooks, C. L. (1993). Statistical clustering techniques for the analysis of long molecular dynamics trajectories: Analysis of 2.2-ns trajectories of YPGDV. *Biochemistry*, *32*(2), 412–420. doi:10.1021/bi00053a005 PMID:8422350

Kim, J. T., Hamilton, A. D., Bailey, C. M., Domaoal, R. A., Domoal, R. A., Wang, L., & Jorgensen, W. L. (2006). FEP-guided selection of bicyclic heterocycles in lead optimization for non-nucleoside inhibitors of HIV-1 reverse transcriptase. *Journal of the American Chemical Society*, *128*(48), 15372–15373. doi:10.1021/ja066472g PMID:17131993

Kitchen, D. B., Decornez, H., Furr, J. R., & Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nature Reviews. Drug Discovery*, *3*(11), 935–949. doi:10.1038/nrd1549 PMID:15520816

Kollman, P. (1993). Free energy calculations: Applications to chemical and biochemical phenomena. *Chemical Reviews*, *93*(7), 2395–2417. doi:10.1021/cr00023a004

Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., & Cheatham, T. E. (2000). Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Accounts of Chemical Research*, *33*(12), 889–897. doi:10.1021/ar000033j PMID:11123888

Krovat, E., Steindl, T., & Langer, T. (2005). Recent Advances in Docking and Scoring. *Current Computer-aided Drug Design*, *1*(1), 93–102. doi:10.2174/1573409052952314

Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R., & Ferrin, T. E. (1982). A geometric approach to macromolecule-ligand interactions. *Journal of Molecular Biology*, *161*(2), 269–288. doi:10.1016/0022-2836(82)90153-X PMID:7154081

Lavecchia, A., & Di Giovanni, C. (2013). Virtual screening strategies in drug discovery: A critical review. *Current Medicinal Chemistry*, *20*(23), 2839–2860. doi:10.2174/09298673113209990001 PMID:23651302

Lin, J.-H., Perryman, A. L., Schames, J. R., & McCammon, J. A. (2002). Computational drug design accommodating receptor flexibility: The relaxed complex scheme. *Journal of the American Chemical Society*, *124*(20), 5632–5633. doi:10.1021/ja0260162 PMID:12010024

Liu, B., Li, S., & Hu, J. (2004). Technological advances in high-throughput screening. *American Journal of Pharmacogenomics: Genomics-Related Research in Drug Development and Clinical Practice*, *4*(4), 263–276. doi:10.2165/00129785-200404040-00006 PMID:15287820

Lobanov, M. Y., Bogatyreva, N. S., & Galzitskaya, O. V. (2008). Radius of gyration as an indicator of protein structure compactness. *Molecular Biology*, *42*(4), 623–628. doi:10.1134/S0026893308040195 PMID:18856071

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Luo, J., & Bruice, T. C. (2002). Ten-nanosecond molecular dynamics simulation of the motions of the horse liver alcohol dehydrogenase*PhCH2O- complex. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(26), 16597–16600. doi:10.1073/pnas.262667599 PMID:12481026

Ma, B., Kumar, S., Tsai, C. J., & Nussinov, R. (1999). Folding funnels and binding mechanisms. *Protein Engineering*, *12*(9), 713–720. doi:10.1093/protein/12.9.713 PMID:10506280

McCammon, J. A., Gelin, B. R., & Karplus, M. (1977). Dynamics of folded proteins. *Nature*, *267*(5612), 585–590. doi:10.1038/267585a0 PMID:301613

Monod, J., Wyman, J., & Changeux, J. P. (1965). On the nature of allosteric transitions: A plausible model. *Journal of Molecular Biology*, *12*(1), 88–118. doi:10.1016/S0022-2836(65)80285-6 PMID:14343300

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., & Schulten, K. (2005). Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, *26*(16), 1781–1802. doi:10.1002/jcc.20289 PMID:16222654

Sinko, W., Lindert, S., & McCammon, J. A. (2013). Accounting for Receptor Flexibility and Enhanced Sampling Methods in Computer-Aided Drug Design: Accounting for Receptor Flexibility. *Chemical Biology & Drug Design*, *81*(1), 41–49. doi:10.1111/cbdd.12051 PMID:23253130

Skjaerven, L., Martinez, A., & Reuter, N. (2011). Principal component and normal mode analysis of proteins; a quantitative comparison using the GroEL subunit. *Proteins*, *79*(1), 232–243. doi:10.1002/prot.22875 PMID:21058295

Sliwoski, G., Kothiwale, S., Meiler, J., & Lowe, E. W. (2013). Computational Methods in Drug Discovery. *Pharmacological Reviews*, *66*(1), 334–395. doi:10.1124/pr.112.007336 PMID:24381236

Talele, T. T., Khedkar, S. A., & Rigby, A. C. (2010). Successful applications of computer aided drug discovery: Moving drugs from concept to the clinic. *Current Topics in Medicinal Chemistry*, *10*(1), 127–141. doi:10.2174/156802610790232251 PMID:19929824

Tripathi, S., Kumar, A., Kumar, B. S., Negi, A. S., & Sharma, A. (2015). Structural investigations into the binding mode of novel neolignans Cmp10 and Cmp19 microtubule stabilizers by in silico molecular docking, molecular dynamics, and binding free energy calculations. *Journal of Biomolecular Structure & Dynamics*, 1–9. doi:10.1080/07391102.2015.1074941 PMID:26212016

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. C. (2005). GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry*, *26*(16), 1701–1718. doi:10.1002/jcc.20291 PMID:16211538

Van Drie, J. H. (2007). Computer-aided drug design: The next 20 years. *Journal of Computer-Aided Molecular Design*, *21*(10-11), 591–601. doi:10.1007/s10822-007-9142-y PMID:17989929

Van Lipzig, M. M. H., terLaak, A. M., Jongejan, A., Vermeulen, N. P. E., Wamelink, M., Geerke, D., & Meerman, J. H. N. (2004). Prediction of ligand binding affinity and orientation of xenoestrogens to the estrogen receptor by molecular dynamics simulations and the linear interaction energy method. *Journal of Medicinal Chemistry*, *47*(4), 1018–1030. doi:10.1021/jm0309607 PMID:14761204

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Mackerell, A. D. (2009). CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry*. http://doi.org/<ALIGNMENT.qj></ALIGNMENT>10.1002/jcc.21367

Vijayakrishnan, R. (2009). Structure-based drug design and modern medicine. *Journal of Postgraduate Medicine*, *55*(4), 301–304. doi:10.4103/0022-3859.58943 PMID:20083886

Villoutreix, B. O., Renault, N., Lagorce, D., Sperandio, O., Montes, M., & Miteva, M. A. (2007). Free resources to assist structure-based virtual ligand screening experiments. *Current Protein & Peptide Science*, *8*(4), 381–411. doi:10.2174/138920307781369391 PMID:17696871

Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., & Case, D. A. (2004). Development and testing of a general amber force field. *Journal of Computational Chemistry*, *25*(9), 1157–1174. doi:10.1002/jcc.20035 PMID:15116359

This work was previously published in Methods and Algorithms for Molecular Docking-Based Drug Design and Discovery edited by Siavoush Dastmalchi, Maryam Hamzeh-Mivehroud, and Babak Sokouti, pages 320-338, copyright year 2016 by Medical Information Science Reference (an imprint of IGI Global).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 33
Recent Advancements in Docking Methodologies

Vijay Kumar Srivastav
Shri Govindram Seksaria Institute of Technology and Science, India

Vineet Singh
Shri Govindram Seksaria Institute of Technology and Science, India

Meena Tiwari
Shri Govindram Seksaria Institute of Technology and Science, India

ABSTRACT

Nowadays molecular docking has become an important methodology in CADD (Computer-Aided Drug Design)-assisted drug discovery process. It is an important computational tool widely used to predict binding mode, binding affinity and binding free energy of a protein-ligand complex. The important factors responsible for accurate results in docking studies are correct binding site prediction, use of suitable small-molecule databases, consistent docking pose, high dock score with good MD (Molecular Dynamics), clarity whether the compound is an inhibitor or agonist, etc. However, still there are several limitations which make it difficult to obtain accurate results from docking studies. In this chapter, the main focus is on recent advancements in various aspects of molecular docking such as ligand sampling, protein flexibility, scoring functions, fragment docking, post-processing, docking into homology models and protein-protein docking.

INTRODUCTION

To accelerate the new drug discovery process, search of small molecules which bind to biological targets of interest, is rapidly increasing. Molecular docking is an important tool used in structure-based drug design which facilitates the drug discovery process with the aid of high speed computers. A typical docking program consists of two essential components viz. sampling and scoring. Sampling is generation of ligand orientations in the binding site of a protein, whereas scoring is prediction of its binding fitness. The top ranked orientation with lowest energy score is predicted as binding mode. Docking is

DOI: 10.4018/978-1-5225-0549-5.ch033

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

now routinely used in hit identification, lead optimization and virtual screening. With the development of protein structure prediction techniques and high speed computers, several advancements have been made in molecular docking methodologies. However, several limitations still exist, which make CADD approaches less reliable than experimental approaches. Several research articles can be found on docking studies but only few of them have succeeded in the development of drug candidate of interest.

This chapter covers general introduction to molecular docking process, its components, limitations and advancements in docking methodologies including ligand sampling, scoring functions, protein flexibility, fragment docking, post-processing, docking into homology models and protein-protein docking along with examples.

BACKGROUND

Enzyme-substrate, drug-protein, drug-nucleic acid, protein-protein and protein-nucleic acid interactions play an important role in many important cellular processes, such as signal transduction, cell regulation, gene expression, enzyme inhibition, antigen-antibody reaction, assembly of multi-domain proteins etc. These interactions form stable protein-protein or protein-ligand complexes which are essential for various biological functions. But, it is quite difficult and expensive to determine the complex structure of proteins by experimental methods i.e. X-ray crystallography and NMR (nuclear magnetic resonance). Thus, structure-based drug design approaches especially docking methodologies are now being considered as an important approach for understanding the protein-protein as well as protein-ligand interactions (Halperin, Ma, Wolfson, & Nussinov, 2002).

In the beginning, structure-based drug design was not easy due to low computational speed and non-availability of 3D (three dimensional) structure of proteins. One key methodology "docking" was pioneered during early 1980s, and became a thrust area of research. The first docking program DOCK was developed in 1982 by Irwin "Tack" Kuntz's group, which considered only geometric complementarities but not the ligand interactions (Kuntz, Blaney, Oatley, Langridge, & Ferrin, 1982). Subsequently, in 1985 Goodford developed a program GRID to understand and predict the ligand interactions in the binding site (Goodford, 1985). Further advancements in CADD resulted in development of various important computational approaches including high-throughput docking and virtual screening (VS). High-throughput docking is primarily used when only the structure of a target and its binding site is available while VS is used to identify a subset of chemical compounds from small-molecules database that contain relatively many hits against the target (Schneider & Böhm, 2002). Till date, several improvements have been made in the docking algorithm, and at the moment more than 50 docking programs are available. Now it is used as a tool not only for hit identification and lead optimization but also for understanding the drug metabolism as well as toxicity prediction (Halgren, 2007).

Development in structure determination techniques like high throughput X-ray crystallography, NMR, genomics and proteomics have increased the number of proteins with known 3D structure and are easily available as a repository of protein structures i.e. PDB (protein data bank). The 3D structure of a protein is an initial and prime requirement of docking program to estimate the binding mode and affinity of interacting ligands (Blundell, Jhoti, & Abell, 2002). Advancement in the structure-based drug design approaches has discovered a number of important drugs like inhibitors of HIV-1 protease, HIV-1 integrase, carbonic anhydrase, neuraminidase, protein kinases etc. Undoubtedly, docking methodologies

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

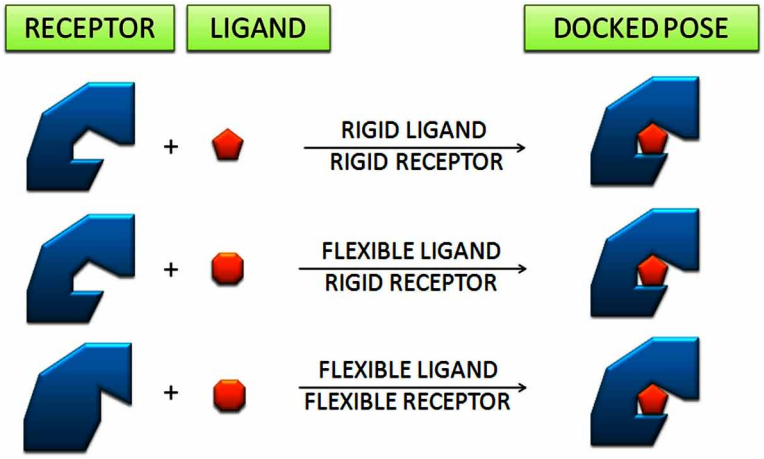
are very important scientific advances to understand the nature of chemical molecules (i.e. drug and receptor), as noted particularly when three top computational scientists won the 2013 Nobel Prize in chemistry. Although, docking methodology has wide applications in drug discovery, there are various problems afflicting docking results and need to be resolved like correct selection of scoring functions, handling of water molecules and metals, covalent interactions and flexibility in the ligand and receptor.

MOLECULAR DOCKING

The complexity of docking methodologies can be represented in the following order: (i) rigid body docking (in which both the receptor and ligand are rigid); (ii) flexible ligand docking (where receptor is rigid, but the ligand is flexible); and (iii) flexible docking (in which both receptor and ligand are flexible). Rigid receptor/flexible ligand model is most commonly used docking methodology (Figure 1). Flexible docking may not be a good choice without a very powerful and high-speed computer to screen more than 10,000 compounds from a database. By contrast, to dock only few compounds in the specific protein binding site, at a specific pH and solvation, flexible docking program might be a good choice. Selection of docking programs, therefore, depends on type of hardware and size of database being screened (Hernández-Santoyo, Tenorio-Barajas, Altuzar, Vivanco-Cid, & Mendoza-Barrera, 2013; Teodoro, Phillips, & Kavraki, 2001).

Recognition of molecular features and their modifications, responsible for specific biological processes are very complex issues that are difficult to be simulated on a computer. A simple molecule contains many conformational degrees of freedom and sampling these degrees of freedom must be carried out with sufficient accuracy to identify the correct orientation that matches the structure of receptor. Ligand sampling and scoring are two important goals of a typical docking program. Sampling must be fast enough to assess thousands of molecules in a defined docking run (Kitchen, Decornez, Furr, & Bajorath, 2004). Calculation of the binding energy of ligand interactions is also one of the important aspects of molecular docking. Methods based on quantum mechanics and empirical energy functions are generally used to calculate this energy. The extensively used docking methods employ search algorithms based on Monte

Figure 1. Types of docking methodologies



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Carlo (MC), genetic algorithm (GA), MD and fragment-based approaches. DOCK, FlexX, GOLD, and ICM are used for high throughput docking of a large database of molecules (Dias & de Azevedo Jr, 2008).

In docking studies, energy estimation is usually carried out with various types of scoring functions. The two main features of a good scoring function are selectivity and efficiency (Teodoro, et al., 2001). Selectivity allows scoring function to differentiate between correct and incorrect docked poses while efficiency allows the docking program to run in a reasonable time. Generally, the available scoring functions are based on forcefield (i.e. force-field based scoring functions) and are used to simulate the receptor-ligand interaction energy and internal ligand energy (strain induced by binding). A forcefield is a function of expressing the energy of a system as a sum total of diverse bonded terms (bond angles, bond distances, torsional angles, etc.) and non-bonded terms [electrostatic and vdW (van der Walls) energy]. These terms are utilized for different types of atoms in the molecule by adjusting a series of empirical parameters. Most of the scoring functions are focused on calculating energetics than solvation and entropic effects. Incorporation of solvation and entropic effects in the scoring function is a challenging task, as it makes the process expensive and time consuming. More complex scoring schemes with electrostatic and vdW interactions along with solvation and entropic effects are now being used to further evaluate the pre-selected conformations with the use of high speed computers. Thus, the accuracy of the scoring function is directly related to its computational speed and expense. Now-a-days statistical scoring functions are also being used that are derived using experimental data. However, there are several problems associated with scoring functions which make docking methodologies complex. In addition, other complications like low resolution of crystallographic protein targets, protein flexibility, induced fit or other conformational changes, and the involvement of water molecules in protein-ligand interactions also exist (Taylor, Jewsbury, & Essex, 2002).

ADVANCEMENTS IN DOCKING METHODOLOGIES

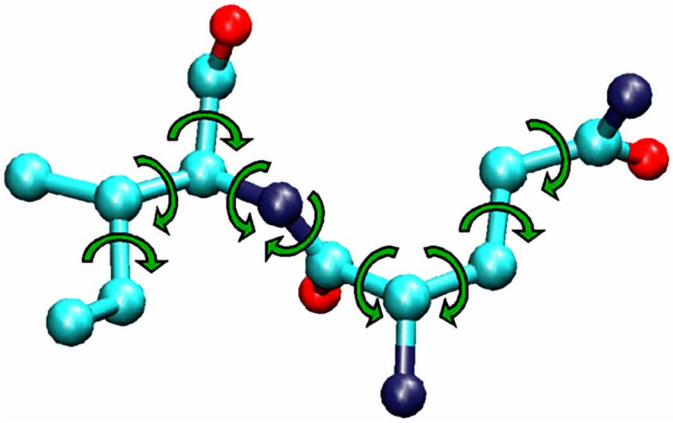
The lock and key theory was used in earliest reported protein-ligand docking methodologies in which both ligand and protein were treated as rigid bodies. Further development introduced docking of flexible ligands to rigid proteins which remains the most popular docking methodology. Recent research is finding the ways of handling the situations where both ligand and protein are flexible. The failure or relative success of the early structure-based drug design methods urges in-depth analysis of docking methodologies to find out the ways of accurate prediction of binding mode and energy. Docking involves a compromise between the docking algorithm speed and its accuracy. In an attempt to achieve higher accuracy, one can employ more advanced scoring functions or more exhaustive sampling of the possible binding modes and flexibility, but these modifications usually add to the computational cost. This section summarizes the recent advancements in docking methodologies and related phenomena.

1. Molecular Representation

The representation of ligand and protein is very important to understand the docking process. Appropriate representation of molecules is now become crucial in analysis of small molecule databases. A molecule can be expressed by a pair of A (collection of atoms) and B (collection of bonds between pairs of atoms). Each atom has some standard properties, such as its vdW radius. There are three main properties associated with each bond: (i) bond length; (ii) bond angle; (iii) number of rotatable bonds (Figure 2).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 2. A drug molecule represented by ball and sticks model
(Teodoro, et al., 2001)



In common practice bond length and angles are considered as fixed because they do not change. Thus, degree of freedom of a molecule mainly depends upon rotatable bonds. Generally, ligands have 3-15 rotatable bonds, while receptors have 1,000-2,000 rotatable bonds (Teodoro, et al., 2001).

In commonly used docking programs, the protomeric and tautomeric states of ligands are user defined. The structure of a ligand can be modified as per the requirement, by adding or removing hydrogens and by adjusting pKa values. Mostly the structures at neutral pH are preferred. The correct atom typing is also very important otherwise wrong definition of acceptor and donor heteroatoms may lead to serious errors in docking results. Example: In an early experiment by Watson and Crick, helical model with purine-pyrimidine hydrogen bonded base pairs failed due to wrong assignment of tautomeric formula (enol forms) for the nucleic acid bases. As the correct tautomeric structures (keto forms) of the bases were assigned, all the key features of the three-dimensional structure of double helical DNA were readily attained (Mohan, Gibbs, Cummings, Jaeger, & DesJarlais, 2005). To describe the stereochemistry of the molecules, it is beneficial to generate all possible diastereoisomers and dock them individually to the receptor. Pipeline Pilot, Stergen and Stereoplex are example of programs available for generating all possible diastereoisomers of a given molecule.

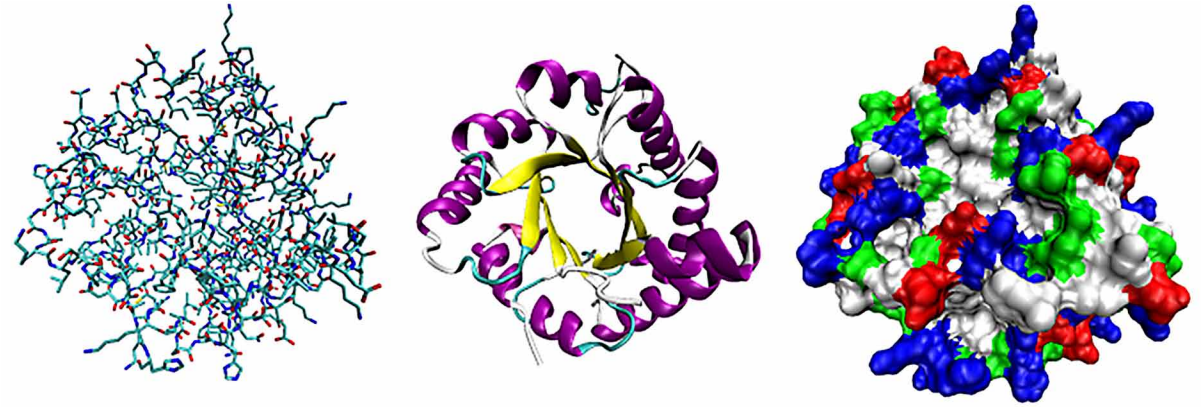
The target protein can be represented by three ways: atomic representation, surface representation and Grid representation. Atomic representation is the most common representation used along with potential energy function and in final ranking procedures. Generally, higher the resolution of a crystal structure, better will be docking result. A docking experiment of reproducing the binding mode using ICM program showed that binding modes of ligands within 1 Å of the bound conformations were better than 2.0 Å. it is also very important that the B-factors of the atoms in the binding site should be reasonable, as high values imply that their coordinates are less reliable (Figure 3). Protein flexibility is described in detail in a later section.

2. Sampling Algorithms

Along with translational, rotational and conformational degrees of freedom, ligand and protein both have a large number of possible binding modes. Thus, it is very difficult to generate all possible conformations.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 3. Representation of 3D structure of the protein triose phosphate isomerase. Left: atom type representation. Middle: backbone conformation, colored by secondary structure. Right: surface representation (acidic residues red, basic residues blue, polar residues green, nonpolar residues white) (http://www.chempep.com/ChemPep-Protein-Overview.htm).



At this stage sampling algorithm plays a very important role. It performs ligand sampling and generates putative ligand poses in the selected binding site. Subsequently the scoring functions select the correct pose based on ranking. In general practice, there are three types of sampling algorithms used in flexible protein-ligand docking: shape matching, systematic search, and stochastic algorithms.

2.1. Shape Matching

The shape matching method places the ligand in the binding site with proper shape compatibility with the binding site of the protein. In this algorithm protein and ligand are considered as pharmacophores and each distance of the pharmacophore within the protein and ligand is calculated for a pair (Meng, Zhang, Mezei, & Cui, 2011). New ligand conformations are regulated by the distance matrix between the pharmacophore and the corresponding ligand atoms. Hydrogen-bond donor and acceptor properties of upcoming conformation are considered during match. Examples of docking programs based on shape matching algorithm include DOCK, EUDOC, FRED, FLOG, LigandFit, MS-DOCK, MDock and Surflex.

2.2. Systematic Search

Systematic search algorithms generally consider all degrees of freedom of ligand to generate all possible binding conformations in the binding site of protein. There are three types of systematic search methods: exhaustive search, fragmentation-based search and conformational ensemble search.

The exhaustive search algorithm operates in docking by rotating all possible rotatable bonds of the ligand in the binding site at a specified interval. Computationally these are too expensive as the search space is very large e.g. Glide and FRED (Kitchen, et al., 2004).

In fragmentation methods, the ligand is introduced in the binding site by dividing it into different fragments. The key fragment which interacts with target protein is selected to dock in the binding site. The remaining fragments are added one by one, to incrementally growing ligand, e.g. ADAM, DOCK,

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

eHiTs, FlexX and LUDI. a new ranking based sampling algorithm have developed within DOCK 6 by softening the vdW interaction energy, improving the performance of the bump filter and by only ranking the layers of ligand growth to manage the sampling algorithm towards the correct pose.

In conformational ensemble methods, ligand flexibility is defined by docking an ensemble of ligand conformations generated programs such as OMEGA. Subsequently, binding modes of ligand from several docking runs are collected and ranked on the basis of their binding energy scores (Roy, Kar, & Das, 2015). Examples of programs based on conformational ensemble are DOCK3.5, FLOG, MS-DOCK, MDock, PhDOCK, and Q-Dock.

2.3. Stochastic Algorithms

Stochastic algorithms are based on the principle of randomness and solve global optimization problems in docking studies. It gives good solutions within reasonable time but do not guarantee the optimum solution, as the random change will be accepted or rejected according to a probabilistic criterion. There are four types of stochastic algorithms: Evolutionary algorithms (EAs), Monte Carlo (MC) methods, Swarm optimization (SO) methods and Tabu search methods (Table 1).

Table 1. Programs based on sampling algorithms

Random/Stochastic	Systematic	Simulation
ICM (MC)	DOCK (incremental)	DOCK
QXP (MC)	FlexX (incremental)	Glide
Prodock (MC)	Glide (incremental)	MOE-Dock
MCDOCK (MC)	Hammerhead (incremental)	Auto Dock
DockVision (MC)	FLOG (database)	Hammerhead
MOE-Dock (MC, Tabu Search)	FRED(exhaustive search)	
Auto Dock (EA, MC)	LUDI	
GOLD (EA)	ADAM	
DIVALI(EA)	eHiTs	
DARWIN (EA)	OMEGA	
MolDock (EA)	PhDOCK	
FLIPDock (EA)	MS-DOCK	
Lead finder (EA)	MDock	
EADock (EA)	Q-Dock	
PSI-DOCK (EA, Tabu Search)		
PRO_LEADS (Tabu Search)		
SODOCK (SO)		
Tribe-PSO (SO)		
PSO@Autodock (SO)		
PLANTS (SO)		

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

2.3.1. Evolutionary Algorithms

EAs searches for the correct binding mode based on evolutionary process in biological systems. The most popular type of EAs is the GA. It is based on Darwin's theory of evolution in which variables are considered as genes (Degrees of freedom of the ligand). These genes make up the 'chromosome' (pose of the ligand). GA searches for poses which show better fitness function through genetic operators like mutation and crossover. The process results in a pose with better fitness function (Michalewicz, 1996; Srivastav & Tiwari, 2015). An improved adaptive GA has been developed which incorporates a multi-population genetic strategy, entropy-based searching with self-adaptation, and the quasi-exact penalty. When this improved adaptive GA was applied to test a dataset of 134 complexes, good pose prediction (<2A˚) was obtained (Kang, Li, Jiang, & Wang, 2009). A hybrid method developed by combining a multi-deme Lamarckinan-GA (LGA) and a gradient-based local optimizer [a modified L-BFGS (Limited-memory-Broyden-Fletcher-Goldfarb-Shanno algorithm)-quasi-Newton method] along with the Gehlhaar scoring function. This hybrid method applied to test a diverse set for flexible ligand-protein docking. It was found faster and superior than other LGAs, which use stochastic optimization methods (Fuhrmann, Rurainski, Lenhof, & Neumann, 2010). GOLD, AutoDock, DIVALI, DARWIN, MolDock, PSI-DOCK, FLIPDock, Lead finder, and EADock are the programs based on EAs.

2.3.2. Monte Carlo Methods

The MC method searches the thermodynamically accessible states of a system (i.e. ligand binding orientation in the target) by generating small random changes in system configuration (i.e. ligand conformation), which can be either accepted or rejected according to the metropolis algorithm. A MC algorithm based method was successfully used to search possible binding modes of molecular fragments of methotrexate (considering translational and rotational degrees of freedom) at a specific site of dihydrofolate reductase from E. ***coli*** (Hart & Read, 1992). The docking programs based on MC methods include DockVision, ICM, QXP, Prodock, and MCDOCK.

2.3.3. Swarm Optimization

Particle Swarm Optimization (PSO) was introduced to solve the conformational sampling problem in flexible protein-ligand docking. It exploits a population of individuals to detect promising regions in the search space. In this context, the population is defined as a swarm, and the individuals are defined as particles (Liu et al., 2013). Examples of docking programs that use SO algorithms include SODOCK, Tribe-PSO, PSO@Autodock, and PLANTS.

2.3.4. Tabu Search

Already explored areas of the conformational space are utilized in tabu search algorithm. An RMSD (root mean square deviation) between the current binding conformation and previously recorded conformation is calculated and used as cutoff to select or reject the upcoming molecular conformation. Only those conformations are accepted which have RMSD value better that the cutoff, otherwise rejected (Westhead, Clark, & Murray, 1997). Examples of tabu search based docking programs are PRO- LEADS and PSI-DOCK.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

2.4. Molecular Dynamics (MD)

Combining docking with MD simulation can handle the flexibility of both ligand and protein more effectively than other algorithms. In order to identify the ligand conformations, MD simulations efficiently use local optimization. An improved MD-based program DynaDock has been developed for docking peptides into flexible receptors (Antes, 2010). This method involves approximate ligand posing, followed by optimized potential MD, which uses soft-core potentials for protein–peptide interactions with a new optimization scheme. It showed better sampling results as compared with conventional MD. The only disadvantage is that it may lead to inadequate sampling when stepping over high energy barriers.

3. Protein Flexibility

A better description of protein-ligand interaction was described by induced-fit model, in which both protein and the ligand are considered as flexible (Koshland, 1958). In this model, to form a minimum energy perfect-fit, both protein and ligand structures change their conformations. Binding of ligand induces conformational changes in protein (induced fit) which ranges from rearrangement of side-chains to large domain motions. Due to the large size and many degrees of freedom of proteins, flexibility became most challenging issue in molecular docking. To overcome this issue several advancements have been done in the docking process such as soft docking, side-chain flexibility, molecular relaxation, and ensemble docking (Table 2). The inclusion of receptor flexibility in docking can result in more accurate binding pose prediction, better virtual screening enrichments, and provide more realistic description of protein-ligand binding interaction.

3.1. Soft Docking

For modeling the protein flexibility, soft docking was the earliest approach used, in which protein is kept rigid but soft scoring functions were used to evaluate the protein-ligand binding. Soft docking works by decreasing the vdW repulsion energy between the binding site and ligand. Decrease in vdW repulsion energy accounts for partial flexibility of protein. It affects minor side-chain movements but not the backbone movements and other structural rearrangements (Ferrari, Wei, Costantino, & Shoichet, 2004; Jiang & Kim, 1991).

3.2. Side-Chain Flexibility

Side-chain flexibility analysis gives valuable insights to improve docking algorithms. In this method the protein backbones are kept fixed and side-chain conformations are sampled to find out the best pose. Amino acid side-chain rotamer libraries are often used to model the conformational space by selecting the preferred side-chain conformations. A dead-end elimination algorithm is used to remove the side-chain conformations which do not contribute in the efficient binding (Desmet, Wilson, Joniau, De Maeyer, & Lasters, 1997; Leach, 1994).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 2. Approaches to incorporate protein flexibility in docking

Approach	Merits	Limitations
Docking with soft potentials	• Low computational cost • Easy to interpret • Can detect subtle changes that other methods overlook	• Only subtle side-chain changes detected • Completely novel conformation cannot be found • Low level of ligand effects • No quantification of conformational changes • Biased by input conformations
Rotamer exploration	• Moderate to high computational cost • depending on size of rotamer library • Can find novel conformation if included in rotamer library • Moderate level of ligand effects • Ideal when active site shows only side-chain rotations	• Only side-chain conformational changes • Multiple solutions may lead to ambiguity in interpretation • May be biased by rotamer library • May fail if desired rotamer not included in library
Multiple protein structures (MPS) multiple dockings	• Side-chain and backbone changes but restricted by ensemble set used • Ligand effects included to moderate level • Multiple ways of generating multiple structures	• High computational cost if many protein structures • Novel conformation cannot be found • Ambiguity in interpretation of multiple equally good solutions • No quantification of conformational changes for specific ligand • Biased by input set of conformations • May fail if specific ligand induces novel conformation not included in MPS set
Multiple protein structures (MPS) ensemble docking	• Side-chain and backbone changes but restricted by ensemble set used • Easy to interpret • Computationally better than multiple structures for VS • Ligand effects included to moderate level	• Moderate computational cost • A small chance of finding novel conformation • No quantification of conformational changes • May fail if ensemble is biased • Average protein structure may not give a real conformation
Current hybrid methods	• Side-chain and backbone movements • Can find novel conformation to some extent • Moderate to high level of ligand effects • Low bias due to input protein conformation	• Moderate to high computational cost • Interpretation may be difficult if multiple equally good solutions found • May miss large conformational changes
MD simulations after docking	• Side-chain and backbone movements but restricted by simulation time • Can find novel conformation to limited extent • Moderate to high level of ligand effects • Quantification of conformational changes • Can handle solvent and ionic effects and membrane-bound proteins	• High computational cost • Analysis of huge output effort-intensive • Biased by starting structure and initial velocity • May fail owing to short times of simulations and may miss best solution

(B-Rao, Subramanian, & Sharma, 2009)

3.3. Molecular Relaxation Process

In the molecular relaxation process first rigid-body docking is performed by placing the ligand into the binding site and minimization of protein-ligand complex is done by using MC, MD, and other methods. This method improves the docking outcome as it is capable of handling both backbone (up to a certain extent) and side-chain flexibility, and thus, preferred over side chain flexibility methods (Apostolakis, Plückthun, & Caflisch, 1998; Davis & Baker, 2009).

3.4. Multiple Protein Structures Ensemble Docking

Docking into multiple receptor conformations is known as ensemble docking and account for receptor flexibility. Instead of using a single rigid protein structure in docking as used in standard docking

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

methods, it involves multiple conformations of the protein to dock the ligand. This ensemble of protein conformations mimics the conformational equilibrium which characterizes the native state of the target protein and provides a structural degree of freedom by which the conformation of the protein model may be matched to fit any particular ligand. Ensemble docking has advantage over other flexible-receptor docking methods like induced-fit docking and capable of handling any type of protein motion (Craig, Essex, & Spiegel, 2010; Knegtel, Kuntz, & Oshiro, 1997; Lang, Ray, Liu, Singh, & Cuny, 2015). FlexE and ICM are examples of ensemble docking programs (Table 3).

3.5. Other Advancements

To handle the protein flexibility, some programs have employed a hybrid approach by combining soft potentials, rotamer exploration and active-site mutation to simulate induced-fit effects, e.g. Schrodinger/Glide (Sherman, Day, Jacobson, Friesner, & Farid, 2006). A knowledge based approach QSiCR (quantitative structure-induced conformation relationship) has been developed to handle protein flexibility. It is similar to QSAR (quantitative structure-activity relationship) but the only difference is, instead of predicting the biological activity it predicts the effect of ligand on structural conformation of key residues in the active site of protein (Subramanian, Sharma, & B-Rao, 2006, 2008). The relaxed complex scheme (RCS) has been developed as a computational methodology by combining the docking algorithm and MD. The merit of this method is that it considers flexibility of both receptor and the docked ligand. Several improvements have been made in RCS, including its extension to virtual screening, more rigorous characterization of local and global binding effects, and improvement in its computational efficiency

Table 3. Recent programs for docking and virtual screening considering protein flexibility

Docking Programs	Novel Features
Autodock	Free open-source EA-based docking program, flexible ligand, and flexible protein side chains.
DOCK	Anchor-and-grow based docking program, flexible ligand, and flexible protein.
GOLD	GA-based docking program, flexible ligand, partial flexibility for protein.
Glide	Exhaustive search-based docking program exists in extra precision, standard precision and virtual high-throughput screening modes, ligand and protein flexible.
SCIGRESS	Desktop/server molecular modeling suite employing linear scaling semi-empirical quantum methods for protein optimization and ligand docking.
ICM	Docking program based on pseudo-brownian sampling and local minimization, ligand and protein flexible.
FlexX, FlexE	Incremental build-based docking program, flexible ligand, protein flexibility through ensemble of protein structure.
Fleksy	Program for flexible and induced fit docking using receptor ensemble (constructed using backbone-dependent rotamer library) to describe protein flexibility.
FITTED	Suite of programs to dock flexible ligands into flexible proteins, relies on a GA to account for flexibility of the two molecules and location of water molecules, and on a novel application of a switching function to retain or displace water molecules and to form potential covalent bonds (covalent docking) with the protein side chains.
Autodock Vina	MC-based docking program, flexible ligand and flexible protein side chains.
FlipDock	GA-based docking program using flextree data structures to represent a protein-ligand complex, flexible ligand, and flexible protein.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

(Amaro, Baron, & McCammon, 2008). A 4D (four dimensional)-docking protocol has been developed for Internal Coordinate Mechanics (ICM), using receptor conformation as a fourth dimension. In this protocol, multiple receptor conformations are represented by multiple grids and used as a variable in global optimization. The method was found to be more accurate as compared to methods using single grid (Abagyan & Totrov, 1994; Bottegoni, Kufareva, Totrov, & Abagyan, 2009).

4. Scoring Functions

The scoring function is an important aspect of protein-ligand docking algorithm, because it directly determines the accuracy and speed of the algorithm and differentiates the correct poses from incorrect poses. The most important issue to scientific community is how to improve the quality of scoring functions? The validation set of GOLD suggested that it is quite difficult to develop scoring functions that can be applied equally well for different protein families, irrespective of their complexity and sophistication. This is just because of uniqueness of binding sites and recognition processes of specific protein-ligand interactions. It was observed that when docking scores are plotted versus logarithm of binding constants, slopes and intercepts are different for structurally distinct binding sites. Thus, adjustment of individual scoring schemes is required on a case by case basis. Scoring functions can be grouped into following categories: force field-based, empirical, knowledge-based, and consensus (Meng, et al., 2011).

4.1. Force Field-Based Scoring Functions

Classical force field scoring functions calculate the binding energy by summing up the individual interaction terms such as vdW energies, electrostatic energies etc. The vdW energies can be described on the basis of Lennard-Jones potential function. The Lennard-Jones potential estimates the acceptability of contact between protein and ligand atoms. The electrostatic energies are assessed by Coulombic calculations. A distance-dependent dielectric function is generally used to describe the contribution of charge-charge interactions, as charge calculations are problematic in modeling the actual protein's environment. Hydrogen bonding, solvation and rotational entropy effects are now being considered in force field-based scoring functions. It has been found that scoring functions which include and explain the solvation and rotational entropy effects are more accurate than standard force-field scoring functions. However, force field-based scoring functions have a disadvantage of slow computational speed (Lyne, 2002).

Docking results with force-field based scoring functions can be refined by using linear interaction energy and free-energy perturbation methods (FEP), which improve the accuracy of binding energy predictions (Meng, et al., 2011). However, these methods together with their simplified approaches such as OWFEG and LIEPROFEC are computationally expensive. The Poisson-Boltzmann/surface area (PB/SA) models and the generalized-Born/surface area (GB/SA) models are typical examples of such implicit solvent models (Morris et al., 1998; Yuriev, Agostino, & Ramsland, 2011; Yuriev & Ramsland, 2013).

4.2. Empirical Scoring Functions

Empirical scoring functions assess the binding energy of a protein-ligand complex by summing up a set of empirical energy components such as hydrogen bonding, ionic interaction, hydrophobic effect, binding entropy term etc. Each component is multiplied by a coefficient and then summed up to give a

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

final score. These Coefficients can be obtained from regression analysis of a test set of ligand-protein complexes with known binding affinities (Eldridge, Murray, Auton, Paolini, & Mee, 1997). GlideScore, PLP, SYBYL/F-Score, LigScore, LUDI, SCORE, X-Score, ChemScore, MedusaScore, AIScore, and SFCscore are examples of empirical scoring functions.

4.3. Knowledge-Based Scoring Functions

Knowledge-based scoring functions employ statistical analysis of structures of protein-ligand complexes to derive the potential parameters such as interatomic contact frequencies and distances between the protein and ligand. It follows the assumption that more favorable an interaction, greater will be the frequency of occurrence. These frequency distributions are then converted into pairwise atom-type potentials. The score is calculated by considering the preferable contacts and penalizing repulsive interactions between each atom in protein and ligand within a given cutoff. It can also be used to screen large chemical compounds databases. It can also deal with some uncommon aromatic-sulphur and cation-π interactions, which are poorly managed by empirical approaches. Knowledge-based scoring functions can be grouped into three categories: traditional atom-randomized reference state, corrected reference state, and circumventing the reference state (S.-Y. Huang & Zou, 2010). The main disadvantage of the atom-randomization approximation is that the effect of excluded volume and interatomic connectivity are overlooked. PMF and DFIRE methods have been introduced as correction terms for the reference state. To circumvent the long-standing reference state problem, ITScore scoring function was derived, which has been tested by prediction of binding mode, binding affinity and virtual screening (Gohlke, Hendlich, & Klebe, 2000). Examples of knowledge-based scoring functions are DrugScore, SMoG, BLEEP, GOLD/ASP, MScore, and KScore.

4.4. Consensus Scoring Functions

Consensus scoring is not a specific type of scoring function but it combines the scoring information from various other scoring functions to find out a correct pose and subsequently balancing the errors of the individual scoring functions. Example: X-CSCORE (which is a combination of ChemScore, DOCK-like, FlexX-like, GOLD-like, and PMF scoring functions). The consensus scoring function based programs work according to the following steps: (i) hierarchical clustering of all poses generated by a docking tool into families represented by a leader; (ii) definition of all consensus pairs from leaders generated by different docking programs; (iii) clustering of consensus pairs into classes, represented by a mean structure; and (iv) ranking the different means starting from the most populated class of consensus pairs (Paul & Rognan, 2002). Other development in consensus scoring method include MultiScore, VoteDock, VLifeDock (uses array of PLPscore, XCscore, steric and electrostatic scores), FLAP-Consensus Docking, ConsDock, PLSDA-DOCET, (Partial Least Squares Discriminant Analysis-Docking Optimized Combined Energetic Terms), a combination of quantitative structure-binding affinity relationship (knowledge based) and MedusaScore (force field-based score), combination of MCSS and GOLD docking with MM/GBSA rescoring, combination of AutoDock4 and AutoDock Vina, combination of DOCK4, FlexX, and PMF, and HarmonyDOCK, etc. (Avram, Pacureanu, Seclaman, Bora, & Kurunczi, 2011; Paul & Rognan, 2002; Plewczynski, Łażniewski, Grotthuss, Rychlewski, & Ginalski, 2011; Poli, Giuntini, Martinelli, & Tuccinardi, 2015).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

4.5. Other Advancements in Scoring Functions

4.5.1. Weighted Score Function

Theoretically, a lower Gibbs free energy of a protein-ligand complex indicates its stability. It has been found that weighted-score function is more accurate than the consensus score (Chen, 2009). Consensus score permits the user to select several or all the scoring functions and evaluate the overall score. Each score function is of equal weight in the consensus scoring function. On the other hand, the weighted-scoring algorithm is based on regression coefficients. A higher coefficient shows a higher weight in this algorithm. Thus, the weighted score is more accurate than the consensus score (Feher, 2006).

$$WS = \sum_{i=1}^{n} Wi \text{ x } \frac{SFi}{SFc} = a \text{ x } pIC_{50} + \text{b} \tag{1}$$

where i is the number of scoring functions; $pIC_{50} = -\log IC_{50}$; SFc is the value of the control scoring function; SFi is the value of the scoring function; Wi is the weight. SFc in Equation I was employed to normalize the weighted equation (Yang, Chang, Chen, & Chen, 2011).

4.5.2. Clustering and Entropy-Based Scoring Methods

To improve the performance of scoring functions, cluster-based scoring methods have been developed which include entropic terms. These methods work by dividing the ligand binding modes into different clusters and entropic effect in each cluster is evaluated by the configurational space of the ligand poses. Combination of these methods with other docking methods improved binding mode prediction (Chang, Belew, Carroll, Olson, & Goodsell, 2008; Ruvinsky, 2007).

4.5.3. Machine-Learning Scoring

Machine learning scoring functions are now being used for prediction of native or near-native pose of a ligand docked to a protein's binding site, and for structure prediction of proteins. In protein structure prediction, machine-learning scores utilize artificial neural network and support vector machine (SVM) to correctly rank the predicted structures of each target according to their quality. It combines multiple features such as physical characteristics and probability distribution based potentials, including different quality structures (Ashtawy & Mahapatra, 2015; L. Li, Wang, & Meroueh, 2011; Qiu, Sheffler, Baker, & Noble, 2008). Examples of machine learning scoring function include SVR (support vector regression)-score, SVR and SVM-KB and SVM-EP scoring functions, and SVM target-specific scoring function.

4.5.4. Combination of Scoring Functions with Experimental (Binding Affinity) Data

An empirical scoring function, PHOENIX, has been developed with 34 descriptors from molecular mechanics calculations, 7 volume-based and shape-based terms to capture entropic contributions, and various surface terms, hydrogen bond donor and acceptor counts, and LogP. The novelty of PHOENIX is that it utilizes experimental data from isothermal titration calorimetry. It results in accurate predictions of binding energy and average prediction of enthalpy and entropy terms.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

4.5.5. MM/GBSA and MM/PBSA Calculations

(MM/PBSA) Molecular Mechanics/Poisson Boltzmann Surface Area and (MM/GBSA) Molecular Mechanics/Generalized Born Surface Area methods calculate binding free energies of macromolecules by combining molecular mechanics and continuum solvation models. Considering the computational efficiency, MM/GBSA can serve as a powerful tool in drug design, where correct ranking of inhibitors is often emphasized (Y. Li, Liu, & Wang, 2010). MM/GBSA and MM/PBSA calculations were applied on 24 diverse protein-ligand complexes, each containing NMR-derived conformations. Modest correlations between experimentally measured binding free energies and the MM/GBSA and MM/PBSA were obtained. A comparison with four selected scoring functions (X-Score, PLP, ChemScore, and DrugScore) showed that MM/GBSA and MM/PBSA are much more computationally demanding (Hou, Wang, Li, & Wang, 2011; Y. Li, et al., 2010).

PROTEIN-PROTEIN DOCKING

PPIs play an important role in many biological events within a cell such as DNA synthesis, gene- expression, signal transduction, immune responses etc. The knowledge of atomic structure of protein-protein complexes is vital for understanding their biological function, for predicting the effect of mutations and for rational design of new PPI-based therapeutic agents. Due to high cost and technical difficulties in structure determination by experimental methods, protein-protein docking is being used for determining the structures of protein-protein complex (Guo, Wisniewski, & Ji, 2014; Sable & Jois, 2015). Various docking programs with varying sampling algorithm and scoring functions have been developed over last decades. As less information about binding site is available in protein-protein docking, the possible binding solutions are obtained by performing global search considering six degrees of freedom (i.e. three translational and three rotational) without requiring theoretical information about the binding site. Based on search strategies, docking programs can be classified into three categories: fast Fourier transform (FFT) or spherical Fourier transform (SFT)-based correlation algorithms, local shape feature matching approaches, and randomized search algorithms (S. Huang, 2015). Examples of docking programs based on FFT correlation search algorithms are DOT, FTDock, GRAMM, MolFit, PIPER, ZDOCK 1.3, ZDOCK 2.1, ZDOCK 2.3.2, ZDOCK 3.0.2, etc. Docking programs ATTRACT, PatchDock, and SwarmDock are based on randomized search algorithm or local shape matching (Table 4).

More information on protein-protein docking programs can be found in several other research papers (Dominguez, Boelens, & Bonvin, 2003; Sable & Jois, 2015; Smith & Sternberg, 2002).

In global docking (or ab initio docking), large number of degrees of freedom in protein's backbone and side chains are treated as rigid bodies to make the search process feasible. To perform the fast evaluation of binding modes, early global protein-protein docking programs like GRAMM and MolFit employed only shape complementarity and hydrophobic interactions, while programs like PatchDock include geometric hashing for fast surface patch matching. To consider full shape complementarity, sophisticated pairwise method has also been included in ZDOCK 2.1. Electrostatic interactions were first considered in FTDock. Since then, electrostatics term is always used along with shape complementarity in upcoming programs like ATTRACT, DOT, HEX, MolFit etc. The desolvation effects have also been included in programs such as FRODOCK, SDOCK, ZDOCK 1.3, and ZDOCK 2.3. Pairwise knowledge-based potentials included in the FFT-based docking programs like PIPER and ZDOCK 3.0 which utilize the

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 4. Protein-protein docking programs

Program	Search Algorithm	Scoring Functions
ATTRACT	Randomized search	Lennard-Jones type effective potentials and electrostatics
DOT	FFT-based correlation	vdW and electrostatic energies
FRODOCK	SFT-based correlation	vdW, electrostatics, and desolvation
FTDock	FFT-based correlation	Shape complementarity and electrostatic interactions
GRAMM	FFT-based correlation	Shape and hydrophobic match
HEX	SFT-based correlation	Surface complementarity and electrostatics
MolFit	FFT-based correlation	Geometric complementarity, hydrophobic complementarity, and electrostatic interactions
PatchDock	Local shape match	Geometric shape complementarity
PIPER	FFT-based correlation	Shape complementarity, electrostatic interactions, and knowledge-based pair potentials
SDOCKb	FFT-based correlation	vdW attractive potential, geometric collision, electrostatic potential, and desolvation energy
ZDOCK 1.3	FFT-based correlation	Shape Complementarity, desolvation, and electrostatics
ZDOCK 2.1	FFT-based correlation	Pairwise shape complementarity
ZDOCK 2.3.2	FFT-based correlation	Pairwise shape complementarity, electrostatics, and desolvation
ZDOCK 3.0.2	FFT-based correlation	Shape complementarity, electrostatics, and knowledge-based pair potentials

(S. Huang, 2015)

information of experimental structures. HADDOCK and RosettaDock programs are used for refinement or site-specific docking with available information about the binding site but the only drawback is high computational cost. CAPRI (Critical Assessment of PRedicted interactions) experiment has been founded to perform the blind predictions of protein-protein docking and scoring algorithms, the results of which contain various biological information (Janin, Wodak, Lensink, & Velankar, 2015). Among different search strategies, FFT (or SFT) correlation methods achieved a good balance between the computational efficiency and the exhaustive global search, in which sampling covers every position of the grid of six-dimensional space (de Vries, Schindler, Chauvot de Beauchêne, & Zacharias, 2015; Schindler, de Vries, & Zacharias, 2015). It is very difficult to study the interactions between membrane proteins that occur inside the membrane bilayers. Recently, a novel method, RosettaMPDock was developed to predict PPIs within the membranes (Leman, Alford, & Gray, 2015).

A Case Study of Hot Spot-Based Design of Protein-Protein Interaction Inhibitor

Identification of hot spots is the most important step in PPI inhibitors design. Docking algorithms are often used to find out the hot spots. In general small molecular probes such as organic solvents are placed on a dense grid around the protein. Using rigid body docking with fast Fourier transform correlation approach, each probe generates many bound poses on the protein surface. For each probe, billions of docked poses are sampled. With the help of RMSD and consensus clustering, the minimized poses are clustered, and hot spots are defined. The binding site having largest cluster is defined as one hot spot. PPI complexes such as Bcl-xL, IL-2IL2R, p53-MDM2, and TNFα-TNFR1 were tested by hot-spot method for druggability. (Kenneth Morrow & Zhang, 2012).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

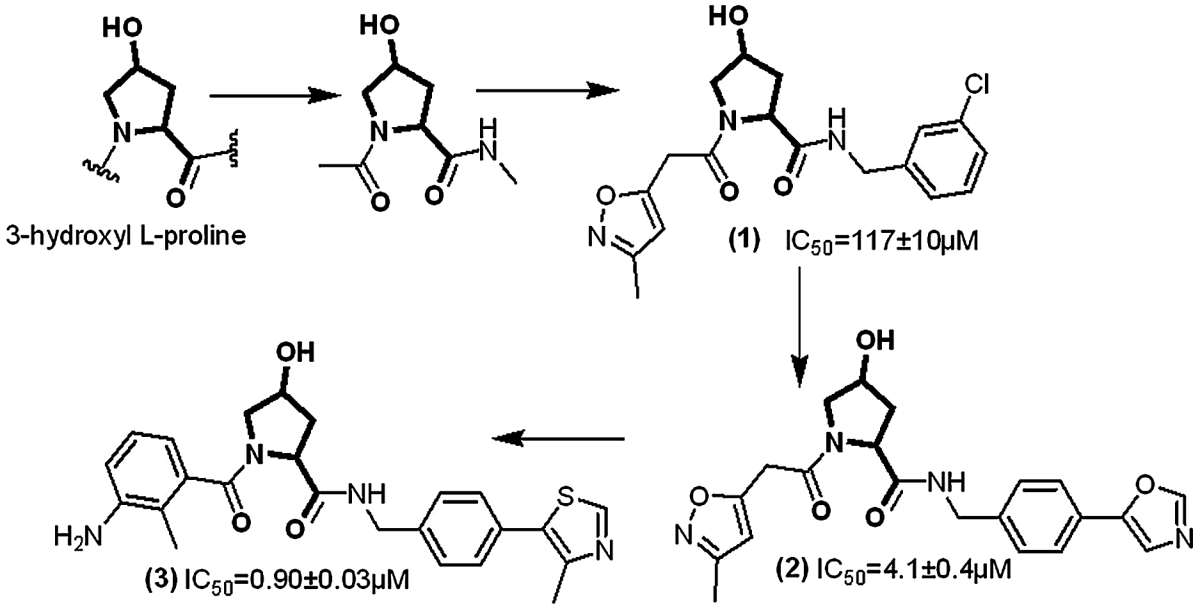
The VHL (Von Hippel-Lindau tumor suppressor) protein is a component of the E3 ligase which interacts with hypoxia-inducible factor 1a (HIF1a) and promotes its degradation by activating ubiquitin proteasomal degradation pathway. VHL/HIF1a inhibitors can be used to treat chronic anemia. 3-hydroxyl-L-proline residue of HIF1a was used as a hot spot for interacting with VHL. Following this hot spot as starting point, new inhibitors were designed with the help of de novo design program BOMB. Compound (1) was discovered which exhibited an IC_{50} value of 117 ± 10 μM (Figure 4). Isoxazole moiety of compound (1) was found to interact with water molecule in the VHL/HIF1a complex. Introduction of an oxazole ring in the para position of the benzyl group resulted in the discovery of compound (2) with an IC_{50} value of 4.1 ± 0.4 μM. Further replacement of oxazole ring with 4-methylthiazole ring led to the generation of compound (3) with an IC_{50} value of 0.90 ± 0.03 μM and better interactions with a hydrophobic pocket in VHL (Guo, et al., 2014).

OTHER ADVANCEMENTS IN DOCKING METHODOLOGIES

1. Inverse Docking

Advancement in docking methodologies includes inverse docking, which has several applications in the field of drug research. First introduced in 2001, inverse docking has been used to identify and validate new potential target of known active compounds. It is also used to distinguish homology models of receptors, prediction of pharmacological profile of a new drug and helps to select less harmful treatment for a disease (Vasseur et al., 2015).

Figure 4. Designing of hydroxyl-proline containing VHL/HIF1a PPI inhibitors
(Guo, et al., 2014)



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

In a study, MDock program with scoring function (ITScore) was used to find out the targets of PRIMA 1. PRIMA 1 is a small molecule which restores mutant p53's tumor suppression function and leads to apoptosis of several cancer cells. Inverse docking involves a PDTD (potential drug target database) from which the target of known compound can be determined (Gao et al., 2008). A molecular surface was defined for each protein in the database, along with sphere points showing initial positions of ligand atoms. Ligand atoms matching the defined sphere points were sampled and ranked by the ITScore. The highest-score conformation was selected by performing rigid body docking of each pair of protein/PRIMA-1-conformations. Then each protein was ranked according to their ITScore. On the basis of sequence identity, the docked structures were then clustered into groups. BLAST search was run on PDTD proteins. Top scored human proteins were then identified and inhibitors were selected as anti-cancer agents for assay (Figure 5). As a result of study, oxidosqualene cyclase was identified as a potential target and Ro 48–8071 as a potent inhibitor of oxidosqualene cyclase and showed significant reduction in viability of cancer cells in bioassay (Grinter, Liang, Huang, Hyder, & Zou, 2011; S. Huang & Zou, 2006).

2. Fragment-Based Docking

Fragment-based drug design approaches are now being frequently used by computational chemists to discover drugs from small molecular fragments. To screen the chemical libraries, FBS (fragment-based screening) is now preferred over HTS (high throughput screening), as it produces higher hit rates, better ligand efficiency and has the ability to sample diverse set of molecules. In Figure 6 advantages of FBS over HTS is represented. It shows that HTS can identify a fully assembled, "drug-sized" ligands but with multiple compromised, non-optimal binding interactions whereas FBS can identify the ligands (with increase ligand affinity) for individual sub-pockets with good binding interactions (Erlanson & Jahnke, 2006).

In the same way, fragment-based docking is based on the fact that each molecular subunit involves at least one favorable interaction in the binding site. The fragment-based docking utilizes the following steps (1) finding optimal position of small molecular fragment in the binding site, (2) linking the small molecular fragments together to form complete molecule that is complementary to the target, and (3) estimating the binding affinity of the resulting molecules. Generally there are four phases of molecular decomposition i.e. ring identification, initial fragment definition, functional group merging, completion of valences. DAIM (Decomposition and Identification of Molecules)-SEED (Solvation Energy for Exhaustive Docking)-FFLD (Fast Flexible Ligand Docking) package and eHiTS are the examples of Fragment based docking programs (Rees, Congreve, Murray, & Carr, 2004).

Several advancements have been made in developing fragment-specific techniques. SERAPHiC (Selected Fragment-Protein Complexes), has been developed as a publicly available dataset of high quality protein-fragment complexes to use it as a validation tool in fragment docking (Favia, Bottegoni, Nobeli, Bisignano, & Cavalli, 2011).

3. Docking in Homology Models

Homology modeling or comparative modeling predicts the 3D-structure of an unknown protein by using homologous sequences of a known protein. The known structure serves as a template and the unknown structure is treated as a target. Homology modeling can be done in following steps: (1) template recogni-

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 5. Inverse docking approach in target identification
(Grinter, et al., 2011)

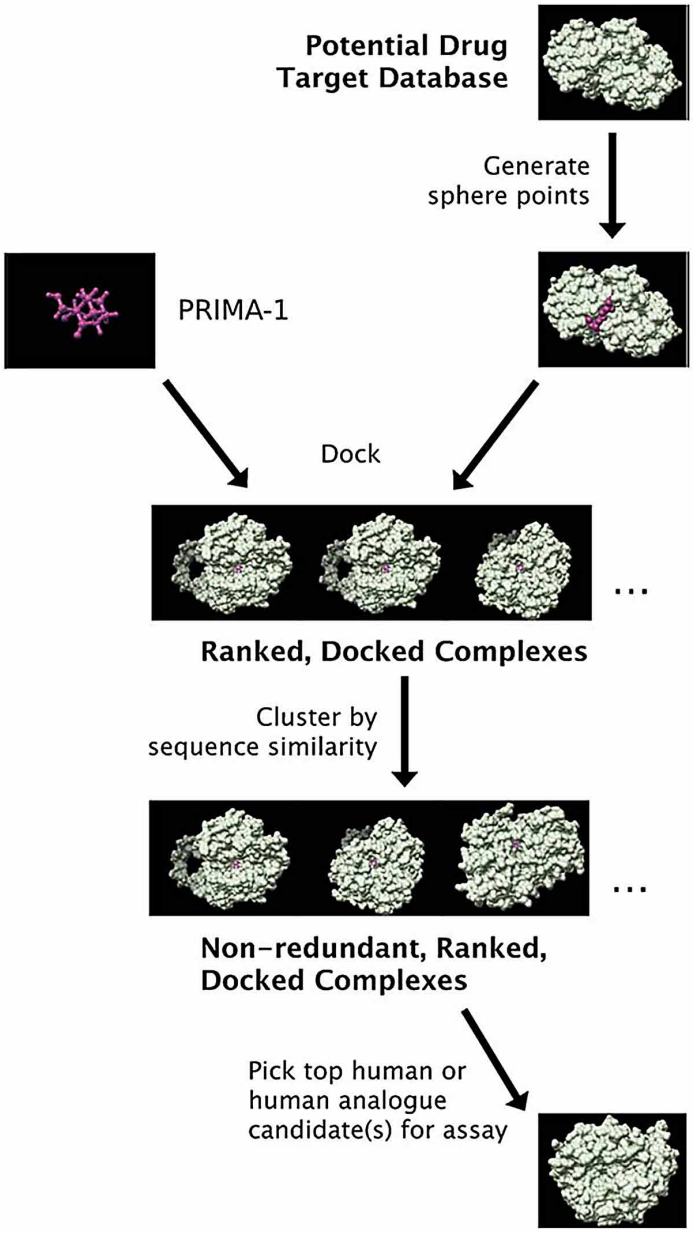
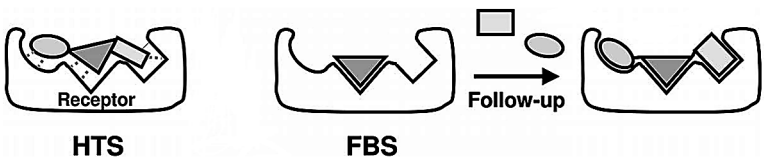


Figure 6. Representation of drawback of HTS (left) and advantages of FBS (right)
(Erlanson & Jahnke, 2006)



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

tion and initial alignment, (2) alignment correction, (3) backbone generation, (4) loop modeling, (5) side-chain modeling, (6) model optimization, and (7) model validation (Krieger, Nabuurs, & Vriend, 2003).

A study described dopamine D1 receptor-agonist interactions by using homology model and mutagenesis. Dopamine D1 receptor is a G protein-coupled receptor which regulates intracellular signaling via agonist activation. Although the number of X-ray structures of G-protein coupled receptors has been steadily increasing, still no structure of the D1 receptor exists. A homology model was constructed for dopamine D1 receptor and effect of various catecholamines was studied. Several key interactions were found with conserved, trans-membrane residues and extra-cellular loop residues. MD simulations were used to examine the interactions. A mutagenesis of 12 orthosteric vicinity residues was used to examine the function of prototypical orthosteric D1 agonists and partial agonists. Mutation studies revealed that the ligands typically have weaker activity at many of the mutant receptors as compared to wild-type D1 receptor, reflecting the loss of interaction between the ligand and the receptor with the mutated side-chain (Mente, Guilmette, Salafia, & Gray, 2015).

Another homology model of CCR2 receptor was presented on the basis of the crystal structure of human beta-2 adrenoceptor (PDB ID: 2RH1). The reported model showed that 99.3% residues were in the core and allowed regions of the ramachandran plot but there was no residue present in the disallowed regions. The model also showed good RMSD value and alignment score with the template, human b2-adrenoceptor. The binding site is found within the transmembrane (TM) region and is large enough for docking the known CCR2 ligands. The docking results validated the homology model to meet various criteria that are necessary in molecular modeling studies (Singh & Elizabeth Sobhia, 2011). Promodel ("http://www.vlifesciences.com/products/BioPredicta/Product_BioPredicta.php," 2015), FoldX, RaptorX, MODELLER, Prime, LOMETS etc., are some examples of homology modeling programs (Chen, 2015).

4. Post-Processing

It is generally found that the top-ranked docking poses do not always represent the correct binding mode and misleading results were found after correlating docking score with biological data. Thus to improve the docking accuracy, post-processing approaches are now being frequently used. Rescoring by MM/PBSA and MM/GBSA techniques is an example of this approach. Using post-processing, binding free energy (MM/PBSA) of 222 Wee1 kinase inhibitors was successfully estimated and derived validated models were used for virtual screening. Prior to post processing, top-ranked docked poses showed wrong binding mode, whereas after post-processing they showed better binding mode. Subsequently, the docked poses giving the lowest binding free energy were selected and these values were used to establish a correlation with experimental data. A significant correlation between ΔG_{cal} and ΔG_{exp} ($r^2 = 0.58$) was obtained, where ΔG is binding energy. The study suggested that post-processing is a very useful approach in refining the docking outcome and can be used for correct prediction of biological activity of novel hits (Fukunishi, 2010; Wichapong, Lawson, Pianwanit, Kokpol, & Sippl, 2010).

5. Study of Absorption, Distribution, Metabolism, and Excretion (ADME) Properties

To reduce the use of experimental animals in clinical studies, research community is now being shifted towards the use of computational approaches which can correctly predict ADME properties of drugs. Now-a-days, docking methodologies are also being frequently used for *in silico* evaluation of ADME.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Cytochrome P450 (CYP450) is an important enzyme which plays an important role in drug metabolism. Several cytochrome P450 isoforms are available which perform different metabolic functions. Various structure-based approaches have been developed which can predict the properties of compounds metabolized by or inhibit P450s (Park & Harris, 2003; Venhorst et al., 2003). Homology models have been developed for human P450 isoforms which serve as templates for docking to predict drug metabolism (de Groot, Ackland, Horne, Alex, & Jones, 1999). A crystal structure of human CYP2C9 complexed with warfarin was determined. It was found that the inhibitor binds to the iron-porphyrin system in the enzyme but do not interact directly with the cofactor. The binding mode of warfarin suggested that CYP2C9 may undergo an allosteric mechanism during its function. The newly discovered binding pocket also suggests that CYP2C9 may simultaneously accommodate multiple ligands during its biological function, and provides a possible molecular basis for understanding complex drug-drug interactions. Thus, metabolism of a compound may be predicted in silico with correct use of docking, MD, and density functional theoretical calculations (Williams et al., 2003).

FUTURE RESEARCH DIRECTIONS

Molecular docking has played a very important role in many successful drug discovery processes. At one hand various advancements have been done in various aspects of docking methodologies while on the other hand scientific community is also looking for refinement of docking results by addressing the limitations. Despite various advancements incorrect binding site prediction, use of undesirable small-molecule databases, inconsistent docking pose, high dock score, failure in MD simulations, lack of clarity whether the compound is an inhibitor or agonist, etc. are still not correctly answered by docking methodologies and urge further research. Even high scored conformations fail in the bioassays showing that there are shortcomings in docking methodologies which rank it lower than experimental approaches. It is still a question that how a correct forcefield can be applied to different set of compounds as well as proteins. Reliability of a protein structure available for docking study is still doubtful. Appropriate validation methods of protein structures are still a prime requirement for accurate docking e.g. Ramachandran plot. Handling large number of degrees of freedom with sufficient accuracy is very complex problem for a docking algorithm and this area still requires further advancement. Therefore, protein and ligand structure validation, protein flexibility, selection of best scoring function for different set of compounds and proteins, docking refinement, correct pose generation, Protein-protein docking, etc. are still an emerging area of future research.

CONCLUSION

The present chapter summarizes the key concepts of docking methodologies, including the recent advances to overcome the challenging issues like protein flexibility, ligand sampling, scoring function, protein-protein docking, etc. In addition, various future directions were also addressed in this chapter.

In the context of target discovery, an example has been discussed in which anti-cancer target identification has been done by inverse docking methodology. Homology modeling also plays an important role in discovery of important compounds for which 3D structure of protein targets is unavailable or

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

difficult to be identified. An example of dopamine receptor and CCR2 receptor is presented here with the help of derived homology model.

From a thermodynamic point of view, scoring functions are still very primitive and need more attention. It has also been found that even high ranked conformations fail in bioassay, post-processing with the inclusion of correct pose sampling and rescoring can improve the docking results. Metabolism of a drug can be predicted with the help of docking studies including the MD and density function calculations, as exemplified by CYP450. As solvation and entropy calculations are costly and require high speed computers, now various scoring function incorporate these terms and give correct results.

Undoubtedly, molecular docking is an important methodology now being used by almost all major pharmaceutical companies and academic institutions. But it is also true that docking seems to have reached a performance plateau and is waiting for a real and important breakthrough.

REFERENCES

Abagyan, R., & Totrov, M. (1994). Biased Probability Monte Carlo Conformational Searches and Electrostatic Calculations for Peptides and Proteins. *Journal of Molecular Biology*, *235*(3), 983–1002. doi:10.1006/jmbi.1994.1052 PMID:8289329

Amaro, R., Baron, R., & McCammon, J. A. (2008). An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *Journal of Computer-Aided Molecular Design*, *22*(9), 693–705. doi:10.1007/s10822-007-9159-2 PMID:18196463

Antes, I. (2010). DynaDock: A new molecular dynamics-based algorithm for protein–peptide docking including receptor flexibility. *Proteins: Structure, Function, and Bioinformatics*, *78*(5), 1084–1104. doi:10.1002/prot.22629 PMID:20017216

Apostolakis, J., Plückthun, A., & Caflisch, A. (1998). Docking small ligands in flexible binding sites. *Journal of Computational Chemistry*, *19*(1), 21–37. doi:10.1002/(SICI)1096-987X(19980115)19:1<21::AID-JCC2>3.0.CO;2-0

Ashtawy, H. M., & Mahapatra, N. R. (2015). Machine-learning scoring functions for identifying native poses of ligands docked to known and novel proteins. *BMC Bioinformatics*, *16*(Suppl 6), S3. doi:10.1186/1471-2105-16-S6-S3 PMID:25916860

Avram, S., Pacureanu, L. M., Seclaman, E., Bora, A., & Kurunczi, L. (2011). PLS-DA - Docking Optimized Combined Energetic Terms (PLSDA-DOCET) Protocol: A Brief Evaluation. *Journal of Chemical Information and Modeling*, *51*(12), 3169–3179. doi:10.1021/ci2002268 PMID:22066983

B-Rao, C., Subramanian, J., & Sharma, S. D. (2009). Managing protein flexibility in docking and its applications. *Drug Discovery Today*, *14*(7–8), 394–400. PMID:19185058

Blundell, T. L., Jhoti, H., & Abell, C. (2002). High-throughput crystallography for lead discovery in drug design. *Nat Rev Drug Discov, 1*(1), 45-54.

Bottegoni, G., Kufareva, I., Totrov, M., & Abagyan, R. (2009). Four-Dimensional Docking: A Fast and Accurate Account of Discrete Receptor Flexibility in Ligand Docking. *Journal of Medicinal Chemistry*, *52*(2), 397–406. doi:10.1021/jm8009958 PMID:19090659

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chang, M. W., Belew, R. K., Carroll, K. S., Olson, A. J., & Goodsell, D. S. (2008). Empirical entropic contributions in computational docking: Evaluation in APS reductase complexes. *Journal of Computational Chemistry*, *29*(11), 1753–1761. doi:10.1002/jcc.20936 PMID:18351616

Chen, Y. (2009). Weighted Equation and Rules-A Novel Concept for Evaluating Protein-Ligand Interaction. *Journal of Biomolecular Structure & Dynamics*, *27*(3), 271–282. doi:10.1080/07391102.2009.10507315 PMID:19795911

Chen, Y. (2015). Beware of docking. *Trends in Pharmacological Sciences*, *36*(2), 78–95. doi:10.1016/j.tips.2014.12.001 PMID:25543280

Craig, I. R., Essex, J. W., & Spiegel, K. (2010). Ensemble Docking into Multiple Crystallographically Derived Protein Structures: An Evaluation Based on the Statistical Analysis of Enrichments. *Journal of Chemical Information and Modeling*, *50*(4), 511–524. doi:10.1021/ci900407c PMID:20222690

Davis, I. W., & Baker, D. (2009). RosettaLigand Docking with Full Ligand and Receptor Flexibility. *Journal of Molecular Biology*, *385*(2), 381–392. doi:10.1016/j.jmb.2008.11.010 PMID:19041878

de Groot, M. J., Ackland, M. J., Horne, V. A., Alex, A. A., & Jones, B. C. (1999). Novel Approach To Predicting P450-Mediated Drug Metabolism: Development of a Combined Protein and Pharmacophore Model for CYP2D6. *Journal of Medicinal Chemistry*, *42*(9), 1515–1524. doi:10.1021/jm981118h PMID:10229622

de Vries, S. J., Schindler, C. E. M., & de Beauchêne, C. (2015). A Web Interface for Easy Flexible Protein-Protein Docking with ATTRACT. *Biophysical Journal*, *108*(3), 462–465. doi:10.1016/j.bpj.2014.12.015 PMID:25650913

Desmet, J., Wilson, I. A., Joniau, M., De Maeyer, M., & Lasters, I. (1997). Computation of the binding of fully flexible peptides to proteins with flexible side chains. *The FASEB Journal*, *11*(2), 164–172. PMID:9039959

Dias, R., & de Azevedo, W. F. Jr. (2008). Molecular Docking Algorithms. *Current Drug Targets*, *9*(12), 1040–1047. doi:10.2174/138945008786949432 PMID:19128213

Dominguez, C., Boelens, R., & Bonvin, A. M. J. J. (2003). HADDOCK: A Protein−Protein Docking Approach Based on Biochemical or Biophysical Information. *Journal of the American Chemical Society*, *125*(7), 1731–1737. doi:10.1021/ja026939x PMID:12580598

Eldridge, M., Murray, C., Auton, T., Paolini, G., & Mee, R. (1997). Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *Journal of Computer-Aided Molecular Design*, *11*(5), 425–445. doi:10.1023/A:1007996124545 PMID:9385547

Erlanson, D. A., & Jahnke, W. (2006). The Concept of Fragment-based Drug Discovery. *Fragment-based approaches in drug discovery*, 1-10.

Favia, A. D., Bottegoni, G., Nobeli, I., Bisignano, P., & Cavalli, A. (2011). SERAPhiC: A Benchmark for in Silico Fragment-Based Drug Design. *Journal of Chemical Information and Modeling*, *51*(11), 2882–2896. doi:10.1021/ci2003363 PMID:21936510

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Feher, M. (2006). Consensus scoring for protein-ligand interactions. *Drug Discovery Today*, *11*(9–10), 421–428. doi:10.1016/j.drudis.2006.03.009 PMID:16635804

Ferrari, A. M., Wei, B. Q., Costantino, L., & Shoichet, B. K. (2004). Soft Docking and Multiple Receptor Conformations in Virtual Screening. *Journal of Medicinal Chemistry*, *47*(21), 5076–5084. doi:10.1021/jm049756p PMID:15456251

Fuhrmann, J., Rurainski, A., Lenhof, H.-P., & Neumann, D. (2010). A new Lamarckian genetic algorithm for flexible ligand-receptor docking. *Journal of Computational Chemistry*, *31*(9), 1911–1918. PMID:20082382

Fukunishi, Y. (2010). Post Processing of Protein-Compound Docking for Fragment-Based Drug Discovery (FBDD): In-Silico Structure-Based Drug Screening and Ligand-Binding Pose Prediction. *Current Topics in Medicinal Chemistry*, *10*(6), 680–694. doi:10.2174/156802610791111452 PMID:20337587

Gao, Z., Li, H., Zhang, H., Liu, X., Kang, L., Luo, X., & Jiang, H. et al. (2008). PDTD: A web-accessible protein database for drug target identification. *BMC Bioinformatics*, *9*(1), 104. doi:10.1186/1471-2105-9-104 PMID:18282303

Gohlke, H., Hendlich, M., & Klebe, G. (2000). Knowledge-based scoring function to predict protein-ligand interactions1. *Journal of Molecular Biology*, *295*(2), 337–356. doi:10.1006/jmbi.1999.3371 PMID:10623530

Goodford, P. J. (1985). A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *Journal of Medicinal Chemistry*, *28*(7), 849–857. doi:10.1021/jm00145a002 PMID:3892003

Grinter, S. Z., Liang, Y., Huang, S.-Y., Hyder, S. M., & Zou, X. (2011). An inverse docking approach for identifying new potential anti-cancer targets. *Journal of Molecular Graphics & Modelling*, *29*(6), 795–799. doi:10.1016/j.jmgm.2011.01.002 PMID:21315634

Guo, W., Wisniewski, J. A., & Ji, H. (2014). Hot spot-based design of small-molecule inhibitors for protein–protein interactions. *Bioorganic & Medicinal Chemistry Letters*, *24*(11), 2546–2554. doi:10.1016/j.bmcl.2014.03.095 PMID:24751445

Halgren, T. (2007). New Method for Fast and Accurate Binding-site Identification and Analysis. *Chemical Biology & Drug Design*, *69*(2), 146–148. doi:10.1111/j.1747-0285.2007.00483.x PMID:17381729

Halperin, I., Ma, B., Wolfson, H., & Nussinov, R. (2002). Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins: Structure, Function, and Bioinformatics*, *47*(4), 409–443. doi:10.1002/prot.10115 PMID:12001221

Hart, T. N., & Read, R. J. (1992). A multiple-start Monte Carlo docking method. *Proteins: Structure, Function, and Bioinformatics*, *13*(3), 206–222. doi:10.1002/prot.340130304 PMID:1603810

Hernández-Santoyo, A., Tenorio-Barajas, A. Y., Altuzar, V., Vivanco-Cid, H., & Mendoza-Barrera, C. (2013). *Protein-Protein and Protein-Ligand Docking*. Academic Press.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Hou, T., Wang, J., Li, Y., & Wang, W. (2011). Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *Journal of Chemical Information and Modeling*, *51*(1), 69–82. doi:10.1021/ci100275a PMID:21117705

Huang, S. (2015). Exploring the potential of global protein-protein docking: An overview and critical assessment of current programs for automatic ab initio docking. *Drug Discovery Today*, *20*(8), 969–977. doi:10.1016/j.drudis.2015.03.007 PMID:25801181

Huang, S., & Zou, X. (2006). An iterative knowledge-based scoring function to predict protein-ligand interactions: I. Derivation of interaction potentials. *Journal of Computational Chemistry*, *27*(15), 1866–1875. doi:10.1002/jcc.20504 PMID:16983673

Huang, S.-Y., & Zou, X. (2010). Advances and Challenges in Protein-Ligand Docking. *International Journal of Molecular Sciences*, *11*(8), 3016–3034. doi:10.3390/ijms11083016 PMID:21152288

Janin, J., Wodak, S. J., Lensink, M. F., & Velankar, S. (2015). Assessing Structural Predictions of Protein–Protein Recognition: The CAPRI Experiment. *Reviews in Computational Chemistry*, *28*, 137–173.

Jiang, F., & Kim, S.-H. (1991). "Soft docking": Matching of molecular surface cubes. *Journal of Molecular Biology*, *219*(1), 79–102. doi:10.1016/0022-2836(91)90859-5 PMID:2023263

Kang, L., Li, H., Jiang, H., & Wang, X. (2009). An improved adaptive genetic algorithm for protein–ligand docking. *Journal of Computer-Aided Molecular Design*, *23*(1), 1–12. doi:10.1007/s10822-008-9232-5 PMID:18777161

Kenneth Morrow, J., & Zhang, S. (2012). Computational Prediction of Protein Hot Spot Residues. *Current Pharmaceutical Design*, *18*(9), 1255–1265. doi:10.2174/138161212799436412 PMID:22316154

Kitchen, D. B., Decornez, H., Furr, J. R., & Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov, 3*(11), 935-949.

Knegtel, R. M. A., Kuntz, I. D., & Oshiro, C. M. (1997). Molecular docking to ensembles of protein structures1. *Journal of Molecular Biology*, *266*(2), 424–440. doi:10.1006/jmbi.1996.0776 PMID:9047373

Koshland, D. E. (1958). Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, *44*(2), 98–104. doi:10.1073/pnas.44.2.98 PMID:16590179

Krieger, E., Nabuurs, S. B., & Vriend, G. (2003). Homology modeling. *Methods of Biochemical Analysis*, *44*, 509–524. PMID:12647402

Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R., & Ferrin, T. E. (1982). A geometric approach to macromolecule-ligand interactions. *Journal of Molecular Biology*, *161*(2), 269–288. doi:10.1016/0022-2836(82)90153-X PMID:7154081

Lang, C. A., Ray, S. S., Liu, M., Singh, A. K., & Cuny, G. D. (2015). Discovery of LRRK2 inhibitors using sequential in silico joint pharmacophore space (JPS) and ensemble docking. *Bioorganic & Medicinal Chemistry Letters*, *25*(13), 2713–2719. doi:10.1016/j.bmcl.2015.04.027 PMID:25998502

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Leach, A. R. (1994). Ligand docking to proteins with discrete side-chain flexibility. *Journal of Molecular Biology*, *235*(1), 345–356. doi:10.1016/S0022-2836(05)80038-5 PMID:8289255

Leman, J. K., Alford, R. F., & Gray, J. J. (2015). Rosetta-MPDock: A Novel Computational Tool for Protein-Protein Docking within the Membrane Bilayer. *Biophysical Journal*, *108*(2), 250a. doi:10.1016/j.bpj.2014.11.1382

Li, L., Wang, B., & Meroueh, S. O. (2011). Support Vector Regression Scoring of Receptor–Ligand Complexes for Rank-Ordering and Virtual Screening of Chemical Libraries. *Journal of Chemical Information and Modeling*, *51*(9), 2132–2138. doi:10.1021/ci200078f PMID:21728360

Li, Y., Liu, Z., & Wang, R. (2010). Test MM-PB/SA on True Conformational Ensembles of Protein-Ligand Complexes. *Journal of Chemical Information and Modeling*, *50*(9), 1682–1692. doi:10.1021/ci100036a PMID:20695488

Liu, Y., Zhao, L., Li, W., Zhao, D., Song, M., & Yang, Y. (2013). FIPSDock: A new molecular docking technique driven by fully informed swarm optimization algorithm. *Journal of Computational Chemistry*, *34*(1), 67–75. doi:10.1002/jcc.23108 PMID:22961860

Lyne, P. D. (2002). Structure-based virtual screening: An overview. *Drug Discovery Today*, *7*(20), 1047–1055. doi:10.1016/S1359-6446(02)02483-2 PMID:12546894

Meng, X.-Y., Zhang, H.-X., Mezei, M., & Cui, M. (2011). Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. *Current Computer-aided Drug Design*, *7*(2), 146–157. doi:10.2174/157340911795677602 PMID:21534921

Mente, S., Guilmette, E., Salafia, M., & Gray, D. (2015). Dopamine D1 receptor-agonist interactions: A mutagenesis and homology modeling study. *Bioorganic & Medicinal Chemistry Letters*, *25*(10), 2106–2111. doi:10.1016/j.bmcl.2015.03.079 PMID:25881819

Michalewicz, Z. (1996). *GAs: What Are They? Genetic Algorithms + Data Structures = Evolution Programs* (pp. 13–31). Springer Berlin Heidelberg. doi:10.1007/978-3-662-03315-9_2

Mohan, V., Gibbs, A. C., Cummings, M. D., Jaeger, E. P., & DesJarlais, R. L. (2005). Docking: Successes and Challenges. *Current Pharmaceutical Design*, *11*(3), 323–333. doi:10.2174/1381612053382106 PMID:15723628

Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., & Olson, A. J. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry*, *19*(14), 1639–1662. doi:10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B

Park, J.-Y., & Harris, D. (2003). Construction and Assessment of Models of CYP2E1: Predictions of Metabolism from Docking, Molecular Dynamics, and Density Functional Theoretical Calculations. *Journal of Medicinal Chemistry*, *46*(9), 1645–1660. doi:10.1021/jm020538a PMID:12699383

Paul, N., & Rognan, D. (2002). ConsDock: A new program for the consensus analysis of protein–ligand interactions. *Proteins: Structure, Function, and Bioinformatics*, *47*(4), 521–533. doi:10.1002/prot.10119 PMID:12001231

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Plewczynski, D., Łażniewski, M., Grotthuss, M. V., Rychlewski, L., & Ginalski, K. (2011). VoteDock: Consensus docking method for prediction of protein–ligand interactions. *Journal of Computational Chemistry*, *32*(4), 568–581. doi:10.1002/jcc.21642 PMID:20812324

Poli, G., Giuntini, N., Martinelli, A., & Tuccinardi, T. (2015). Application of a FLAP-Consensus Docking Mixed Strategy for the Identification of New Fatty Acid Amide Hydrolase Inhibitors. *Journal of Chemical Information and Modeling*, *55*(3), 667–675. doi:10.1021/ci5006806 PMID:25746133

Qiu, J., Sheffler, W., Baker, D., & Noble, W. S. (2008). Ranking predicted protein structures with support vector regression. *Proteins: Structure, Function, and Bioinformatics*, *71*(3), 1175–1182. doi:10.1002/prot.21809 PMID:18004754

Rees, D. C., Congreve, M., Murray, C. W., & Carr, R. (2004). Fragment-based lead discovery. *Nat Rev Drug Discov, 3*(8), 660-672.

Roy, K., Kar, S., & Das, R. N. (2015). Other Related Techniques. In K. R. K. N. Das (Ed.), *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment* (pp. 357–425). Boston: Academic Press.

Ruvinsky, A. (2007). Calculations of protein-ligand binding entropy of relative and overall molecular motions. *Journal of Computer-Aided Molecular Design*, *21*(7), 361–370. doi:10.1007/s10822-007-9116-0 PMID:17503189

Sable, R., & Jois, S. (2015). Surfing the Protein-Protein Interaction Surface Using Docking Methods: Application to the Design of PPI Inhibitors. *Molecules (Basel, Switzerland)*, *20*(6), 11569–11603. doi:10.3390/molecules200611569 PMID:26111183

Schindler, C. E. M., de Vries, S. J., & Zacharias, M. (2015). iATTRACT: Simultaneous global and local interface optimization for protein–protein docking refinement. *Proteins: Structure, Function, and Bioinformatics*, *83*(2), 248–258. doi:10.1002/prot.24728 PMID:25402278

Schneider, G., & Böhm, H.-J. (2002). Virtual screening and fast automated docking methods. *Drug Discovery Today*, *7*(1), 64–70. doi:10.1016/S1359-6446(01)02091-8 PMID:11790605

Sherman, W., Day, T., Jacobson, M. P., Friesner, R. A., & Farid, R. (2006). Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects. *Journal of Medicinal Chemistry*, *49*(2), 534–553. doi:10.1021/jm050540c PMID:16420040

Singh, R., & Elizabeth Sobhia, M. (2011). Homology modeling of human CCR2 receptor. *Medicinal Chemistry Research*, *20*(9), 1704–1712. doi:10.1007/s00044-010-9497-9

Smith, G. R., & Sternberg, M. J. E. (2002). Prediction of protein–protein interactions by docking methods. *Current Opinion in Structural Biology*, *12*(1), 28–35. doi:10.1016/S0959-440X(02)00285-3 PMID:11839486

Srivastav, V. K., & Tiwari, M. (2015). k-nearest neighbor molecular field analysis based 3D-QSAR and in silico ADME/T studies of cinnamoyl derivatives as HIV-1 integrase inhibitors. *Medicinal Chemistry Research*, *24*(2), 684–700. doi:10.1007/s00044-014-1183-x

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Subramanian, J., Sharma, S., & B-Rao, C. (2006). A Novel Computational Analysis of Ligand-Induced Conformational Changes in the ATP Binding Sites of Cyclin Dependent Kinases. *Journal of Medicinal Chemistry*, *49*(18), 5434–5441. doi:10.1021/jm060172s PMID:16942017

Subramanian, J., Sharma, S., & B-Rao, C. (2008). Modeling and Selection of Flexible Proteins for Structure-Based Drug Design: Backbone and Side Chain Movements in p38 MAPK. *ChemMedChem*, *3*(2), 336–344. doi:10.1002/cmdc.200700255 PMID:18081134

Taylor, R. D., Jewsbury, P. J., & Essex, J. W. (2002). A review of protein-small molecule docking methods. *Journal of Computer-Aided Molecular Design*, *16*(3), 151–166. doi:10.1023/A:1020155510718 PMID:12363215

Teodoro, M. L., Phillips, G. N., & Kavraki, L. E. (2001). *Molecular docking: A problem with thousands of degrees of freedom.* Paper presented at the Robotics and Automation, 2001. doi:10.1109/ROBOT.2001.932674

Vasseur, R., Baud, S., Steffenel, L. A., Vigouroux, X., Martiny, L., Krajecki, M., & Dauchez, M. (2015). Inverse docking method for new proteins targets identification: A parallel approach. *Parallel Computing*, *42*, 48–59. doi:10.1016/j.parco.2014.09.008

Venhorst, J., ter Laak, A. M., Commandeur, J. N. M., Funae, Y., Hiroi, T., & Vermeulen, N. P. E. (2003). Homology Modeling of Rat and Human Cytochrome P450 2D (CYP2D) Isoforms and Computational Rationalization of Experimental Ligand-Binding Specificities. *Journal of Medicinal Chemistry*, *46*(1), 74–86. doi:10.1021/jm0209578 PMID:12502361

Westhead, D. R., Clark, D. E., & Murray, C. W. (1997). A comparison of heuristic search algorithms for molecular docking. *Journal of Computer-Aided Molecular Design*, *11*(3), 209–228. doi:10.1023/A:1007934310264 PMID:9263849

Wichapong, K., Lawson, M., Pianwanit, S., Kokpol, S., & Sippl, W. (2010). Postprocessing of Protein−Ligand Docking Poses Using Linear Response MM-PB/SA: Application to Wee1 Kinase Inhibitors. *Journal of Chemical Information and Modeling*, *50*(9), 1574–1588. doi:10.1021/ci1002153 PMID:20712342

Williams, P. A., Cosme, J., Ward, A., Angove, H. C., Matak Vinkovic, D., & Jhoti, H. (2003). Crystal structure of human cytochrome P450 2C9 with bound warfarin. *Nature, 424*(6947), 464-468.

Yang, S., Chang, S., Chen, H., & Chen, Y. (2011). Identification of Potent EGFR Inhibitors from TCM Database@Taiwan. *PLoS Computational Biology*, *7*(10), e1002189. doi:10.1371/journal.pcbi.1002189 PMID:22022246

Yuriev, E., Agostino, M., & Ramsland, P. A. (2011). Challenges and advances in computational docking: 2009 in review. *Journal of Molecular Recognition*, *24*(2), 149–164. doi:10.1002/jmr.1077 PMID:21360606

Yuriev, E., & Ramsland, P. A. (2013). Latest developments in molecular docking: 2010–2011 in review. *Journal of Molecular Recognition*, *26*(5), 215–239. doi:10.1002/jmr.2266 PMID:23526775

This work was previously published in Methods and Algorithms for Molecular Docking-Based Drug Design and Discovery edited by Siavoush Dastmalchi, Maryam Hamzeh-Mivehroud, and Babak Sokouti, pages 267-294, copyright year 2016 by Medical Information Science Reference (an imprint of IGI Global).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 34
Methods for Docking and Drug Designing

Ahmad Abu Turab Naqvi
Jamia Millia Islamia, India

Md. Imtaiyaz Hassan
Jamia Millia Islamia, India

ABSTRACT

Molecular docking is the prediction of conformational complementarity between ligand and receptor molecule. The process of docking integrates two schematic approaches namely sampling of ligand conformations and ranking of selected conformations based on scoring functions. The authors have discussed established methodologies for molecular docking and well-known tools implementing these methods. A brief account of different classes of scoring functions such as force field based, empirical, knowledge based, and descriptor based scoring functions is given along with the exemplary implementations of these scoring functions. By replacing test and trial based ligand screening with structure based virtual screening, molecular docking has helped in shortening the duration of novel drug discovery up to some extent. However, the developments made in the field of drug discovery are assisted by the advances in the techniques of molecular docking, but there is strong need of enrichment in the techniques, especially in scoring functions, to tackle the inbound problems of de novo drug discovery.

INTRODUCTION

All biological processes and molecular functions are carried out by diverse classes of macromolecules such as nucleic acids, proteins, carbohydrates, etc. Sometimes macromolecules perform certain activity on their own, but certain functions are carried out due to the intermolecular interactions in which the activity of a molecular species is induced by other molecules and vice versa. The concept of molecular docking is based on the phenomenon of intermolecular interactions between leading molecular species that play a key role in the working mechanisms of various biological processes and molecular functions. Basically, the event of docking upholds the binding of one molecule into the conformational space of another molecule leading to a specific activity or function. Docking can be seen between two proteins

DOI: 10.4018/978-1-5225-0549-5.ch034

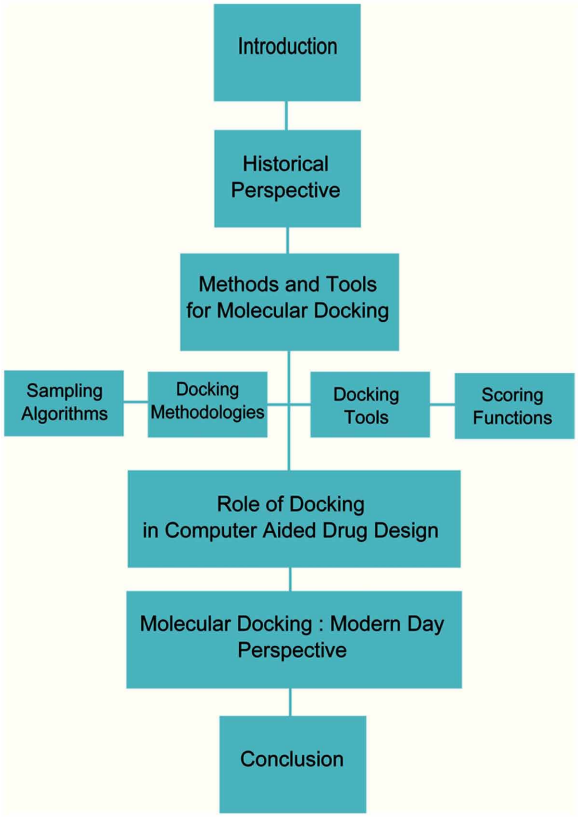
Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

having conformational complementariness such as protein- protein docking, or between one small molecule (usually called ligand) and a protein, DNA-protein docking, etc. In the context of computer-aided drug design, ligand- protein docking holds significant status. Therefore, in the upcoming paragraphs the authors will objectively discuss various aspects of ligand- protein docking in particular and molecular docking in general.

Several reviews and evaluation reports have been published in recent years covering important aspects of molecular docking, recent advances in methods of docking, evolution of search algorithms, developments in scoring functions, and advancements in docking tools (Ferreira, Dos Santos, Oliva, & Andricopulo, 2015; Kitchen, Decornez, Furr, & Bajorath, 2004; Meng, Zhang, Mezei, & Cui, 2011; G. M. Morris & Lim-Wilby, 2008; Oshiro, Kuntz, & Dixon, 1995; Verdonk, Cole, Hartshorn, Murray, & Taylor, 2003). In this chapter, the authors have tried to bring forward the objective study of small molecule-protein docking in the context of computer-aided drug design including brief idea of various preceding events of molecular docking such as target identification, lead optimization, and virtual screening. This chapter further elaborates the historical perspective of molecular docking, role of docking in computer aided drug design, an introduction to various methods and established tools for molecular docking and underlying scoring functions that assess the quality of conformational complementarity between ligand and receptor molecule. In this way, the flow of the chapter (Figure 1) goes with an introduction and historical background of molecular docking along the brief description of docking methodologies, docking

Figure 1. Flow chart showing covered topics of the chapter



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

tools, search algorithms and scoring function. In context of drug discovery, the authors have discussed the role of molecular docking in drug discovery by discussing some case studies. The authors also have discussed modern perspective of molecular modeling including the present day challenges of the area with respect to its involvement in drug discovery.

HISTORICAL PERSPECTIVE

Though, in *post genomic* era molecular docking has met revolutionary developments making it a more advanced concept aided by other fields of study such as computer science, physics, and statistics, etc., but the traces of the basic idea of docking way back to "Lock and Key Hypothesis" proposed by Emil Fischer in 1894 in the context of enzyme – substrate geometric interaction. During the development phases of molecular biology, several modifications have been proposed to understand the inbound complexity of docking. In 1958, Koshland came forward with the idea of "induced fit'' in his modification of "Lock and Key Hypothesis". He proposed that the conformational changes in the active site of the receptor are induced by the binding of the substrate (Koshland, 1958).

Before the period of late 1970s and 80s, molecular docking was studied more as a phenomenon of molecular interaction between the molecules having binding affinity. The idea of docking as a distinguished concept originates in the works of late 1970s. The term docking also has its origin in the studies conducted in late 70s (Wodak & Janin, 1978). The developments made in the 70s were more or less confined to interactions between a large molecule such as protein- protein docking (Wodak & Janin, 1978). More polished idea of docking of small molecules with proteins emerged in the early 1980s in the work of Kuntz et al., (Kuntz, Blaney, Oatley, Langridge, & Ferrin, 1982).

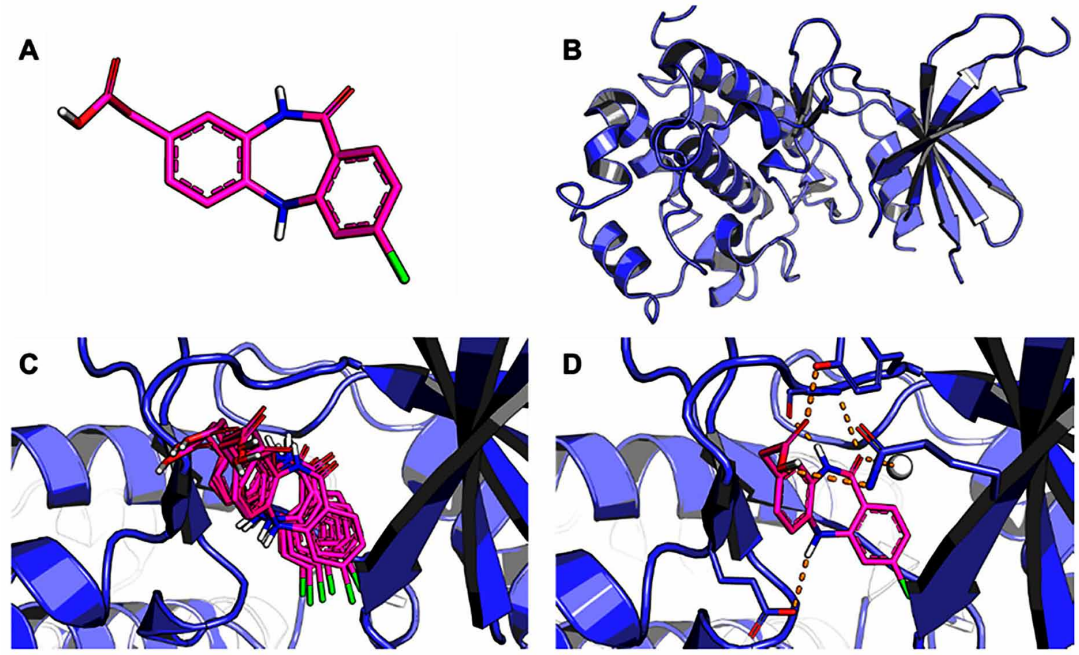
The field of molecule docking as a distinguished field of study has been developed parallel to the developments in computer science and technology. Molecular docking emerged significantly in the context of computer-aided drug design hand in hand with several techniques like target identification, virtual screening or high throughput screening, molecular dynamics simulation, scoring functions, search algorithms, etc. (Kitchen, et al., 2004) (Meng, et al., 2011). Going through all the processes of development, ever since late 1970s, molecular docking has emerged as more developed field of study which integrates a huge variety of entities such drug discovery databases, ligand libraries, tools of molecular docking and various tools of evaluation.

METHODS AND TOOLS FOR MOLECULAR DOCKING

In this section, the authors will discuss some established methods of molecular docking with a brief introduction to some well developed tools based on these methodologies. Technically, docking is defined as the prediction of conformational complimentary between ligand and receptor molecule (proteins in the context of CADD) followed by the application of scoring function for the assessment of binding in terms of binding energy (G. M. Morris & Lim-Wilby, 2008). The whole process of docking integrates two schematic approaches i.e., sampling of ligand conformations and ranking of selected conformations based on scoring functions (Meng, et al., 2011). A typical example of outline of docking process is shown in Figure 2.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 2. Figure showing (A) three-dimensional structure of the ligand; (B) three-dimensional structure of the receptor; (C) the ligand is docked into the binding cavity of the receptor and the putative conformations are explored; (D) the most likely binding conformation and the corresponding intermolecular interactions are identified.
Figure was taken from the Ferreira et al. 2015.



Sampling Algorithms

Sampling of ligand conformations in the active site of the receptor is done using various sampling algorithms. During the last decades, there is a revolutionary development in sampling algorithms. All of these sampling algorithms ranging from probabilistic to stochastic have their merits and demerits based on the computational cost and time efficiency. On the other hand, these algorithms are chosen on the priority basis depending upon the size of involved molecules. For example, molecular dynamics simulation has comparatively higher computational cost therefore it is usually preferred for small molecules.

Brooijmans and Kuntz (Brooijmans & Kuntz, 2003) in their review of sampling algorithms of molecular docking have proposed basic categories of searching algorithms based on the approach applied by these search algorithms. The basic categories of search algorithms are systematic, stochastic and simulation based (Kitchen, et al., 2004). In this way, incremental construction (Rarey, Kramer, Lengauer, & Klebe, 1996) and database search strategies are kept in systemic search categories. In incremental construction method, ligand is broken into fragments and the fragment showing higher complementarity to the active site conformation is docked inside the active site pocket as the 'anchor' then the remaining fragments of the ligand are docked in an incremental manner. Monte Carlo (Hart & Read, 1992) and Genetic algorithms (Jones, Willett, Glen, Leach, & Taylor, 1997; Garrett M. Morris et al., 1998) are the most established methods for stochastic searching of conformational space. Monte Carlo produces different conformations of ligand by inducing bond rotation and rigid body translation (Meng, et al.,

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

2011). Genetic algorithms are the second most used search algorithms in molecular docking. The concept of Genetic algorithms is based on Darwin's evolution theory (Oshiro, et al., 1995). Genetic algorithms carry out ligand conformation search in term of genes, chromosomes and mutations. Energy based assessment of ligand structure is done and favoured conformation is taken to the next step based on the theory of 'survival of the fittest'. The ability of these methods to modify ligand conformation makes them more preferred to other search methods. Molecular dynamics simulation represents the simulation based methods of conformation search it is the most preferred approach in simulation ever since its development (Brooijmans & Kuntz, 2003).

Docking Methodologies

The differentiation between docking methodologies evolves on the basis of flexibility/rigidity of the ligand and/or receptor molecule. Thus, there are three basic classes of docking methods based the above-mentioned criteria:-

- **Rigid Ligand and Rigid Receptor Docking:** In this approach both ligand and receptor molecules are considered rigid. The relative rigidity is due to the lesser number of degrees of for rotational and translational space.
- **Flexible Ligand and Rigid Receptor Docking:** This is the most implemented docking methodology in almost all the leading docking software. The concept of flexible ligand rigid receptor docking is based on the 'induced fit' prototype of ligand- receptor interaction.
- **Flexible Ligand and Flexible Receptor Docking:** Ideally, docking in its traditional approach deals with ligand flexibility having rigid receptor in the frame. The idea of receptor flexibility is so far a challenging task in molecular docking (Lexa & Carlson, 2012). However, molecular dynamics simulation techniques have efficiency of handling the problem of receptor flexibility, but due to the high computation cost such approaches are less preferred.

Docking Tools

A vast range of docking tools has been developed by incorporating above mentioned search algorithms and docking methodologies so far such as AutoDock, GOLD, DOCK, Glide, FRED, FITTED and CDocker etc. Table 1 provides a brief description of some of the leading docking tools and web based docking servers including the information about implemented search algorithms and scoring functions.

Scoring Functions

In structure based virtual screening, the prediction of possible interaction between ligand and receptor in feasible conformational space is assessed in terms of binding energy. The elimination of 'false' ligand poses from the population of 'true' ones on the basis of minimal binding free energy is carried out using different classes of scoring functions. These scoring functions, on the basis of the fact that they consider for the estimation of binding energy, are divided into different classes such force field based scoring functions, empirical scoring functions, knowledge based scoring functions, scoring based on consensus (Kitchen, et al., 2004) and descriptor based scoring functions (Liu & Wang, 2015). Recent advances in molecular docking were mainly focused on strengthening the scoring strategies by including certain

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 1. List of docking tools and web servers

S. No.	Software	Description	Platform
21.	Autodock 4.2.6	It is based on Monte Carlo search algorithm for flexible ligand search. The scoring functions used in AutoDock are force field based.	Desktop based
22.	DOCK 6.7	It uses incremental construction strategy for docking using AMBER molecular mechanics scoring function with implicit solvent	Desktop based
23.	GOLD	It is a Genetic algorithm based docking software. It uses force field based scoring functions for the evaluation of docking.	Desktop based
24.	Glide	It provides exhaustive search method for ligand screening.	Desktop based
25.	ICM	It is based on pseudo-Brownian sampling and local minimization. Provides ligand and protein flexibility.	Desktop based
26.	FlexX	It is an incremental construction based docking software. It comes with flexible ligand and flexible protein docking strategy.	Desktop based
27.	FITTED	It is a genetic algorithm based docking soft with more explicit features with explicit water molecule, side chain flexibility, displaceable water molecules etc.	Desktop based
28.	AutoDock Vina	It is Monte Carlo based. Provides ligand flexibility and side chain flexibility for receptor.	Desktop based
29.	FRED	It performs a systematic and non stochastic search for ligand conformations. It uses Chemgauss4 scoring function for the assessment of the search	Desktop based
30.	CDocker	It is a CHARMm based docking platform. It generates random ligand conformations using molecular dynamics.	Desktop based
31.	MOE suite 2014.9	Provides diverse range of modules used in molecular docking and conformations evaluation.	Desktop based
32.	ADAM	It is an Automated docking tool. Suitable for virtual high throughput screening.	Desktop based
33.	SwissDock	Suitable for predicting molecular interaction between small molecule and target.	Web based
34.	DockingServer	Online docking interface of docking.	Web based
35.	1-Click Docking	Online docking platform with WebGL/Javascript based molecule viewer of GLmol.	Web based
36.	Pardock	Provides all-atom energy based Monte Carlo, rigid protein ligand docking and parallel processing mode.	Web based
37.	PatchDock	It provides structure prediction of protein-protein and protein-small molecule complexes using shape complementarity principles.	Web based
38.	Dock blaster	Online server for structure based ligand discovery.	Web based
39.	PLATINUM	Calculates ligand- receptor interaction on the basis of Molecular Hydrophobicity Potential" (MHP).	Web based
40.	iScreen	Cloud computing web server for virtual screening and de novo drug design based on TCM database.	Web based

factors such as entropy effects, water molecules, and solvent effects were not considered in previous versions of these scoring functions. Computational time and cost were two main factors that are still to be addressed. The assessment of docking in the stochastic environment is needed to get better results, but the chances of increase in computational time and cost are the major obstacle in this approach. Parallel developments in computational techniques are paving the way for the elimination of this problem. In this section, the authors will discuss the aforementioned classes of scoring functions along with the docking tools that implement these scoring functions.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Force-Field Based Scoring Functions

Force field was conceptualized in the first half of the twentieth century. It remains the most used strategy in a variety of fields including molecular modeling through molecular dynamics simulation, energy minimization strategies, and molecular docking. Force field based scoring functions estimate the binding energy of ligand-receptor complex by considering non bonded interactions. These non bonded interactions include electrostatic interactions and van Waals interactions. The estimation of binding energy as a sum of different molecular mechanics terms is carried in using various functions. Electrostatic interactions are expressed in terms of coulombic formula whereas Lennard - Jones function is used for van der Waals interactions. Force field functions used by diverse docking tools are expressed in various parametric sets of molecular mechanics terms, such as DOCK (Allen et al., 2015) and AutoDock (G. M. Morris et al., 2009) use Amber force filed where as CDocker (Wu, Robertson, Brooks, & Vieth, 2003) uses CHARMm force field scoring functions. Advance versions of DOCK, AutoDock and GOLD have implemented some extensions in force field scoring functions such as hydrogen bonding terms, entropy and solvent factors (Meng, et al., 2011). Despite being widely used in a range of established docking tools, force field based scoring functions have certain limitations. The problem of high computational time leads to slow speed of estimations. To avoid slow speed, cut off distances are used for non bonded interactions. These cutoff distances affect the accuracy of estimation while considering long distance effects of non bonded interactions (Kitchen, et al., 2004; Meng, et al., 2011).

Empirical Scoring Functions

Empirical scoring functions estimate the binding energy of the ligand – receptor complex in terms of various energy terms such as hydrogen bonds, hydrophoboic effect, ionic interactions, binding entropy and rotatable bonds in the ligand (Bohm, 1994; Eldridge, Murray, Auton, Paolini, & Mee, 1997). Empirical scoring functions use the coefficients of the energy terms obtained by the regression analysis of the binding energies of known complex structures. LUDI (Bohm, 1992), Chemscore implemented in GOLD (Verdonk, et al., 2003), SCORE (Wang, Liu, Lai, & Tang, 1998) are the well-known examples of empirical scoring functions.

Knowledge Based Scoring Functions

Knowledge based scoring functions have certain advantages over other scoring functions due to the computational simplicity that they provide. The idea of knowledge based scoring functions is to use the statistical inferences from the binding energy potentials of known ligand – receptor complex structures hence the term knowledge based scoring functions. The frequent iteration of a certain pair wise atomic contact between ligand and receptor is considered energetically favoured. Thus, inferences for the query ligand-receptor complex are made based on these assumptions. SMoG (DeWitte & Shakhnovich, 1996), PMF (Muegge & Martin, 1999), and DrugScore (Gohlke, Hendlich, & Klebe, 2000) are the examples of knowledge based scoring functions.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Descriptor Based Scoring Functions

Jie Liu and Renxiao Wang (Liu & Wang, 2015) in a recent classification of scoring functions discussed this relatively new kind of scoring functions. The idea of descriptor based scoring functions is evolved from the quantitative structure - activity relationship (QSAR) analysis strategy. In the context of CADD, QSAR analysis is used for the evaluation of the interaction between the drug molecule and the target protein (receptor). QSAR equation is expressed using various molecular descriptors as parameters. These descriptors include surface or shape properties, hydrogen bonds, molecular weight, etc. RF-Score (Ballester & Mitchell, 2010), ID- Score (Li, Yang, Wang, Li, & Yang, 2013), and SF-Score (Zilian & Sotriffer, 2013) are the examples of descriptor based scoring functions.

Consensus Scoring

Consensus scoring has remained the preferred way of scoring in various techniques of computer based predictions. Scoring functions have certain limitations in assessing the quality of predictions. To tackle this situation, researchers have found the way of consensus scoring. Ligand-receptor complementarity is assessed using different scoring functions. The structure that gets better scores in maximum number of scoring strategies is accepted. CScore (Clark, Strizhev, Leonard, Blake, & Matthew, 2002) is an example of consensus based scoring functions. It uses DOCK, ChemScore, GOLD, and FlxX scoring functions.

ROLE OF DOCKING IN COMPUTER AIDED DRUG DESIGN

The principal objective of this chapter is to discuss molecular docking focusing its importance in computer aided drug design. During the last two decades, molecular docking has drastically revolutionized the field of drug discovery. Parallel developments in computer aided drug discovery having molecular docking incorporated as a principal phenomenon has made the process of drug discovery more time efficient. The integration of computer based techniques applied to preceding events of docking such as targeted identification, high throughput screening, lead optimization, etc., are now accomplished in less time, hence making the process less time-consuming (Gu, Zhang, & Yuan, 2014; Jorgensen, 2004). In order to understand the role of Molecular Docking and available docking programs, the authors are discussing here some case studies made during recent past in which molecular docking played essential role to bring forward the consequences favouring the possibility of drug discovery against certain diseases.

Design and Synthesis of Piperidine Derivatives as Novel Human Heat Shock Protein 70 Inhibitors for the Treatment of Drug-Resistant Tumors

Yanqun et al. (Zeng, Cao, Zhang, Li, & Zhong, 2015) have carried out designing and synthesis of piperidine derivatives as the inhibitors for human heat shock protein 70 (HSP70). HSP70s are significant due to their involvement in various essential biological processes such as maintenance of protein homeostasis, regulation of apoptosis and involvement in the binding of (Tumour necrosis factor (TNF) and TNF-related apoptosis inducing ligand (TRAIL) to the receptors. The study is mainly focused on the practicality of HSP70s in the treatment of cancers and drug-resistant tumours. HSP70s play a vital role in cancer resistance during the treatment of tumour cells. The phenomenon is backed by the over

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

expression of HSP70s in cancer patients which also favours the spreading of cancerous infection to other cells falling in range of the cancer affected area. The study puts a significant stress on lapatinib-resistant breast cancer cells while performing the cell viability assays. Study suggests that compounds HSP70-36/37/40/43/46 are significant in inhibiting the growth of breast cancer cells. Developments made in the area were motivated by the unavailability of enough inhibiters against HSP70s.

The study is carried out with combination of *in silico* and *in vivo* techniques. In context, the authors will discuss the concerned part, i.e. *in silico* in detail. A set of tools and databases has been used for the purpose. The synthesis of novel inhibitors is based on N-terminal nucleotide binding domain of HSP70. Crystal structure of nucleotide binding domain in complex with adenosine diphosphate (ADP) is used for the screening. In this way, docking domain is detected using MVD4.0 software. Molergo Visual Docking (MVD) is an integrated docking platform. It provides high accuracy prediction of protein-ligan interaction. It is specifically equipped with the ability to predict potential binding sites in the target protein and also binding modes for the ligand. In this case, Cambridge Database and a patented library of 243,712 compounds are used for virtual fragment screening. The calculation of binding modes for the ligands in HSP70 ATPase domain has been performed using Discovery studio. Discovery Studio is an integrated program contained a diverse range of modules to perform various calculations and predictions used in molecular docking. Prediction of binding modes for all 67 compounds has been performed using Discovery Studio. Discovery studio is also equipped with modules to perform energy minimization and other energy related calculations. Energy minimization has also been performed using the modules available in Discovery Studio. To predict the drugability of the compounds ADMET screening was performed using the prescribed modules of Discovery Studio. In this study, an array of different programs for molecular docking, virtual fragment screening, and binding mode prediction has been used to maintain the credibility of the *in silico* approach. In recent time, an affinity for consensus based approach for computer based predictions has been developed in the researchers. It gives the researchers a way of comparing the results provided by different programs and then dragging consensus based consequences. Finally, Surflex-SYBYL6.9 was used for precise docking of the ligands. Novel derivatives of piperidine have good anticancer bioactivities. They also possess very low rate of inhibiting HSP70 ATPase activity. Therefore, newly derived inhibitors of HSP70 may provide ways of inhibiting the activity of HSP70 in cancer treatments.

Discovery of Akt Kinase Inhibitors through Structure-Based Virtual Screening and Their Evaluation as Potential Anticancer Agents

In recent years, several studies have been conducted for discovery and development of potential inhibitors against known target proteins that are involved in disease causing mechanisms. The structure based virtual screening and docking serves the purpose for successful development of such inhibiting compounds. Chuan et al. (Chuang et al., 2015) have conducted a similar study to discovery novel inhibitors against the activity of human Atk kinase using the structure based virtual screening. Akt is an essential regulator in the phosphatidylinositol 3-kinase PI3K/Akt signaling pathway therefore it is considered a potential drug target in cancer therapy. In this study, x-ray crystal structure of human Atk in complex with pyrrolopyrimidine inhibitor is used to perform virtual screening. Virtual screening has been performed using a dataset of 35,367 compounds that is derived from SPECS through Zinc database of chemical compounds. DOCK4.0 has been used for structure based virtual screening. ATP binding site of Atk is selected for screening. Consequently, a set of 48 compounds was selected that are showing inhibiting

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

activity against Atk activity. Biological assessment of these 48 compounds for Atk kinase inhibition provided 26 final compounds that exhibited potential inhibition effect against Atk. Out of these compounds a48 (compounds were named as a1-a48) showed highest inhibition activity against Atk kinase. Finally, selected 26 test compounds were further taken for Cytotoxicity Evaluation on HCT-116 Cancer Cells and HEK-293 Normal Cells. Out of these 26 compounds 12 compounds expressed higher cytotoxicity against HCT-116 Cancer Cells. Eventually, compounds a46 and a48 were selected for further study on the basis of their IC_{50} values against colon cancer cells. Finally selected compounds show promising inhibition activity against Atk. Docking of both these compounds in the ATP binding site of Atk is performed using GOLD5.0 (Fig). GOLD performs genetic algorithm based flexible ligand docking simulation. On the basis of biological evaluation and docking studies, both compounds a46 and a48 are found potential inhibitors of Atk kinase activity that makes them potential anticancer agents.

Development of Effective Drug against H1N1 Pathogen

Mallipeddi et al., (Mallipeddi, Kumar, White, & Webb, 2014) in their review of recent developments in the field of computer aided drug discovery have discussed the applications of molecular docking in the development of anti-influenza drug. The pandemic outbreak of H1N1 influenza in 2009 has affected a majority of human population globally. Drugs effective against the H1N1 pathogen namely Zanamivir and oseltamivir are developed using the advanced techniques of molecular docking and virtual screening.

Development of Anti-HIV Drugs that Inhibit the C-C Chemokine Receptor Type 5 (CCR5) Activity

AIDS, ever since its first case detected in human has remained the one of the severest health issues for human race. Human immunodeficiency virus (HIV) causes AIDS in human beings. It enters the target cells through co receptors. In a review published, Planche & Cordeiro (Speck-Planche & Cordeiro, 2011) have highlighted the role of computer aided drug discovery methods in the development of anti-HIV drugs that inhibit the C-C chemokine receptor type 5 (CCR5) activity. CCR5 receptors are the most crucial co receptors that facilitate the process of HIV infection.

MOLECULAR DOCKING: MODERN DAY PERSPECTIVE

So far in this chapter, the authors have discussed molecular docking, its historical perspective, a brief introduction to various search algorithms, docking methods, docking programs and some case studies in which docking played role towards obtaining the desired results. However, with the advancement in Computation power that more strong computing devices, high performance servers, high power processing units, supercomputers, molecular docking is also undergone fruitful development in recent years, but there are still some challenges that this ever-developing field is facing. These challenges are evolved due to certain limitations and drawbacks of docking methods and programs. Though, on evolutionary level, docking programs have been developed effectively to throw out possible limitations, but the inability of most of the programs to contour the problems evolved due to the more stochastic behaviour of the molecular entities.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

With the possible development of docking tools, in time, field of drug discovery has been revolutionized astronomically. The major concern for the development of docking programs has remained biomolecular flexibility specially flexibility of receptor protein. The other problem that has been managed, though not solved completely, is application of docking strategies for stochastic systems. In drug discovery process, the *in silico* part, i.e. computational predictions, target identification and final binding of ligand in receptor molecular hence identification of potential drug has been controlled up to some extent. The major problem is faced during laboratory experimentation of predicted compounds. This is still a challenge in the field of drug discovery for experimentalists.

Leading docking tools like, AutoDock, AutoDock Vina, GOLD, DOCK, and Glide, etc. have implemented advanced search algorithms and scoring functions to counter the problems of protein flexibility and stochasticity. However, the problem has been dealt up to some level; still there are practical instances where some of these programs fail to serve the purpose. Docking studies for highly stochastic and complex systems such as membrane proteins face various problems. To counter such kinds of difficulties, researchers have started using consensus approaches for molecular docking. For a problem, they use an array of docking and virtual screening programs in place of using single software for everything.

As the authors have discussed above, the major challenges in molecular docking are structural flexibility, stochasticity of the involved system and implementation of computational prediction at experimentation level with a promising precision that may pave the way of development of the drug. Most of the problems and limitations that docking programs face can be solved by advancing the search algorithms and scoring function. There is a need of more stochastic search algorithms such as genetic algorithm and implementation of molecular dynamics simulation with precise parameters. Currently, the parameters used in scoring functions and simulation are knowledge based. The deterministic parameters are less effective in making stochastic predictions. Hence, there is a need of developing the strategies to involve more and more stochastic parameters.

CONCLUSION

Molecular docking has become the principal weapon to conquest the problems of structure based virtual screening. It is in the way of revolutionary progress ever since its conceptualization. The parallel developments in computer technology have aided the progress of molecular docking in various manners. Though, docking has reached possible limits of enrichments up to some extent, but there are still some problems that are to be addressed for further developments in molecular docking. The intricate stochastic behaviour of the biological process is still a major challenge for the researchers. Furthermore, the interaction between ligand and receptor takes place in a more stochastic manner having so many factors affecting the phenomenon such as water molecules, entropy effects, solvent effects, etc. So, there is a strong need of enrichment of docking techniques with more focus on the strength of efficient scoring functions that can tackle the situation in a more realistic manner. There is also an undeniable inability of docking techniques and scoring functions to the problem of *de novo* drug design. Most of the scoring functions and search algorithms use the traditional approach of the parametric approximation on the basis of known experimental data. These approximations have their limits while applied in the process of *de novo* drug design. However, docking techniques become more powerful with efficient scoring functions, but further progress is strongly needed. Scoring functions with more and more stochastic parameters will be helpful in the betterment of *de novo* drug design processes.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

REFERENCES

Allen, W. J., Balius, T. E., Mukherjee, S., Brozell, S. R., Moustakas, D. T., Lang, P. T., & Rizzo, R. C. et al. (2015). DOCK 6: Impact of new features and current docking performance. *Journal of Computational Chemistry*, *36*(15), 1132–1156. doi:10.1002/jcc.23905 PMID:25914306

Ballester, P. J., & Mitchell, J. B. (2010). A machine learning approach to predicting protein-ligand binding affinity with applications to molecular docking. *Bioinformatics (Oxford, England)*, *26*(9), 1169–1175. doi:10.1093/bioinformatics/btq112 PMID:20236947

Bohm, H. J. (1992). LUDI: Rule-based automatic design of new substituents for enzyme inhibitor leads. *Journal of Computer-Aided Molecular Design*, *6*(6), 593–606. doi:10.1007/BF00126217 PMID:1291628

Bohm, H. J. (1994). The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *Journal of Computer-Aided Molecular Design*, *8*(3), 243–256. doi:10.1007/BF00126743 PMID:7964925

Brooijmans, N., & Kuntz, I. D. (2003). Molecular recognition and docking algorithms. *Annual Review of Biophysics and Biomolecular Structure*, *32*(1), 335–373. doi:10.1146/annurev.biophys.32.110601.142532 PMID:12574069

Chuang, C. H., Cheng, T. C., Leu, Y. L., Chuang, K. H., Tzou, S. C., & Chen, C. S. (2015). Discovery of Akt kinase inhibitors through structure-based virtual screening and their evaluation as potential anticancer agents. *International Journal of Molecular Sciences*, *16*(2), 3202–3212. doi:10.3390/ijms16023202 PMID:25648320

Clark, R. D., Strizhev, A., Leonard, J. M., Blake, J. F., & Matthew, J. B. (2002). Consensus scoring for ligand/protein interactions. *Journal of Molecular Graphics & Modelling*, *20*(4), 281–295. doi:10.1016/S1093-3263(01)00125-5 PMID:11858637

DeWitte, R. S., & Shakhnovich, E. I. (1996). SMoG: De Novo Design Method Based on Simple, Fast, and Accurate Free Energy Estimates. 1. Methodology and Supporting Evidence. *Journal of the American Chemical Society*, *118*(47), 11733–11744. doi:10.1021/ja960751u

Eldridge, M. D., Murray, C. W., Auton, T. R., Paolini, G. V., & Mee, R. P. (1997). Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *Journal of Computer-Aided Molecular Design*, *11*(5), 425–445. doi:10.1023/A:1007996124545 PMID:9385547

Ferreira, L. G., Dos Santos, R. N., Oliva, G., & Andricopulo, A. D. (2015). Molecular docking and structure-based drug design strategies. *Molecules (Basel, Switzerland)*, *20*(7), 13384–13421. doi:10.3390/molecules200713384 PMID:26205061

Gohlke, H., Hendlich, M., & Klebe, G. (2000). Knowledge-based scoring function to predict protein-ligand interactions. *Journal of Molecular Biology*, *295*(2), 337–356. doi:10.1006/jmbi.1999.3371 PMID:10623530

Gu, W. G., Zhang, X., & Yuan, J. F. (2014). Anti-HIV drug development through computational methods. *The AAPS Journal*, *16*(4), 674–680. doi:10.1208/s12248-014-9604-9 PMID:24760437

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Hart, T. N., & Read, R. J. (1992). A multiple-start Monte Carlo docking method. *Proteins*, *13*(3), 206–222. doi:10.1002/prot.340130304 PMID:1603810

Jones, G., Willett, P., Glen, R. C., Leach, A. R., & Taylor, R. (1997). Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology*, *267*(3), 727–748. doi:10.1006/jmbi.1996.0897 PMID:9126849

Jorgensen, W. L. (2004). The many roles of computation in drug discovery. *Science*, *303*(5665), 1813–1818. doi:10.1126/science.1096361 PMID:15031495

Kitchen, D. B., Decornez, H., Furr, J. R., & Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nature Reviews. Drug Discovery*, *3*(11), 935–949. doi:10.1038/nrd1549 PMID:15520816

Koshland, D. E. (1958). Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, *44*(2), 98–104. doi:10.1073/pnas.44.2.98 PMID:16590179

Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R., & Ferrin, T. E. (1982). A geometric approach to macromolecule-ligand interactions. *Journal of Molecular Biology*, *161*(2), 269–288. doi:10.1016/0022-2836(82)90153-X PMID:7154081

Lexa, K. W., & Carlson, H. A. (2012). Protein flexibility in docking and surface mapping. *Quarterly Reviews of Biophysics*, *45*(3), 301–343. doi:10.1017/S0033583512000066 PMID:22569329

Li, G. B., Yang, L. L., Wang, W. J., Li, L. L., & Yang, S. Y. (2013). ID-Score: A new empirical scoring function based on a comprehensive set of descriptors related to protein-ligand interactions. *Journal of Chemical Information and Modeling*, *53*(3), 592–600. doi:10.1021/ci300493w PMID:23394072

Liu, J., & Wang, R. (2015). Classification of current scoring functions. *Journal of Chemical Information and Modeling*, *55*(3), 475–482. doi:10.1021/ci500731a PMID:25647463

Mallipeddi, P. L., Kumar, G., White, S. W., & Webb, T. R. (2014). Recent advances in computer-aided drug design as applied to anti-influenza drug discovery. *Current Topics in Medicinal Chemistry*, *14*(16), 1875–1889. doi:10.2174/1568026614666140929153812 PMID:25262796

Meng, X. Y., Zhang, H. X., Mezei, M., & Cui, M. (2011). Molecular docking: A powerful approach for structure-based drug discovery. *Current Computer-aided Drug Design*, *7*(2), 146–157. doi:10.2174/157340911795677602 PMID:21534921

Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., & Olson, A. J. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry*, *19*(14), 1639–1662. doi:10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, *30*(16), 2785–2791. doi:10.1002/jcc.21256 PMID:19399780

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Morris, G. M., & Lim-Wilby, M. (2008). Molecular docking. *Methods in Molecular Biology (Clifton, N.J.)*, *443*, 365–382. doi:10.1007/978-1-59745-177-2_19 PMID:18446297

Muegge, I., & Martin, Y. C. (1999). A general and fast scoring function for protein-ligand interactions: A simplified potential approach. *Journal of Medicinal Chemistry*, *42*(5), 791–804. doi:10.1021/jm980536j PMID:10072678

Oshiro, C. M., Kuntz, I. D., & Dixon, J. S. (1995). Flexible ligand docking using a genetic algorithm. *Journal of Computer-Aided Molecular Design*, *9*(2), 113–130. doi:10.1007/BF00124402 PMID:7608743

Rarey, M., Kramer, B., Lengauer, T., & Klebe, G. (1996). A fast flexible docking method using an incremental construction algorithm. *Journal of Molecular Biology*, *261*(3), 470–489. doi:10.1006/jmbi.1996.0477 PMID:8780787

Speck-Planche, A., & Cordeiro, M. N. (2011). Current drug design of anti-HIV agents through the inhibition of C-C chemokine receptor type 5. *Current Computer-aided Drug Design*, *7*(4), 238–248. doi:10.2174/157340911798260287 PMID:22050678

Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W., & Taylor, R. D. (2003). Improved protein-ligand docking using GOLD. *Proteins*, *52*(4), 609–623. doi:10.1002/prot.10465 PMID:12910460

Wang, R., Liu, L., Lai, L., & Tang, Y. (1998). SCORE: A New Empirical Method for Estimating the Binding Affinity of a Protein-Ligand Complex. *Molecular Modeling Annual, 4*(12), 379-394. doi:10.1007/s008940050096

Wodak, S. J., & Janin, J. (1978). Computer analysis of protein-protein interaction. *Journal of Molecular Biology*, *124*(2), 323–342. doi:10.1016/0022-2836(78)90302-9 PMID:712840

Wu, G., Robertson, D. H., Brooks, C. L. III, & Vieth, M. (2003). Detailed analysis of grid-based molecular docking: A case study of CDOCKER-A CHARMm-based MD docking algorithm. *Journal of Computational Chemistry*, *24*(13), 1549–1562. doi:10.1002/jcc.10306 PMID:12925999

Zeng, Y., Cao, R., Zhang, T., Li, S., & Zhong, W. (2015). Design and synthesis of piperidine derivatives as novel human heat shock protein 70 inhibitors for the treatment of drug-resistant tumors. *European Journal of Medicinal Chemistry*, *97*, 19–31. doi:10.1016/j.ejmech.2015.04.043 PMID:25935385

Zilian, D., & Sotriffer, C. A. (2013). SFCscore(RF): A random forest-based scoring function for improved affinity prediction of protein-ligand complexes. *Journal of Chemical Information and Modeling*, *53*(8), 1923–1933. doi:10.1021/ci400120b PMID:23705795

KEY TERMS AND DEFINITIONS

Complementarity: Complimentary properties of ligand and receptor in docked complex structure.
Docking: Posing of Ligand molecule into the active site pocket of receptor.
Drug: Chemical compound having therapeutic properties against any disease.
Force Field: Function that expresses energy of a system as sum of non bonded interactions.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Lead Identification: Process of identifying suitable chemical compound showing the properties of drug.

Ligand: Small molecule that binds to the receptor.

Receptor: Molecule that provides conformational space for the binding of ligand.

This work was previously published in Methods and Algorithms for Molecular Docking-Based Drug Design and Discovery edited by Siavoush Dastmalchi, Maryam Hamzeh-Mivehroud, and Babak Sokouti, pages 39-53, copyright year 2016 by Medical Information Science Reference (an imprint of IGI Global).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 35
Protein–Ligand Docking Methodologies and Its Application in Drug Discovery

Sanchaita Rajkhowa
Tezpur University, India

Ramesh C. Deka
Tezpur University, India

ABSTRACT

Molecular docking is a key tool in structural biology and computer-assisted drug design. Molecular docking is a method which predicts the preferred orientation of a ligand when bound in an active site to form a stable complex. It is the most common method used as a structure-based drug design. Here, the authors intend to discuss the various types of docking methods and their development and applications in modern drug discovery. The important basic theories such as sampling algorithm and scoring functions have been discussed briefly. The performances of the different available docking software have also been discussed. This chapter also includes some application examples of docking studies in modern drug discovery such as targeted drug delivery using carbon nanotubes, docking of nucleic acids to find the binding modes and a comparative study between high-throughput screening and structure-based virtual screening.

1. INTRODUCTION

Traditionally, the discovery of new drugs has arisen from observations that various plant extracts possess interesting biological effects. However, early users of such plant extracts did not understand or realize which components in the material were responsible for achieving these therapeutic benefits. The main difference between modern and age-old drug discovery is that the identification of the composition of matter, or the active form, within the medicine itself. Drug discovery and development is mostly a complex, expensive, time consuming and market-driven process with very few novel drug candidates actually making it through the Food and Drug Administration (FDA) for approval. Modern drug dis-

DOI: 10.4018/978-1-5225-0549-5.ch035

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

covery process involves rational approach with knowledge about enzymes and receptors and the use of complex synthetic methods and technologies such as combinatorial chemistry, microwave assisted organic synthesis and high-throughput (HTS) biological screening methods (Patil, 2012).

In recent years, there has been an increase in the number of new therapeutic targets for modern drug discovery with enormous contribution to many structural details of proteins and protein-ligand complexes. This has been possible as a result of the completion of the human genome project, along with the development of high-throughput protein purification, crystallography and nuclear magnetic resonance spectroscopy. Experimental high-throughput screening (HTS) and combinatorial chemistry were developed in order to overcome the various problems faced by lead discovery in drug development. Some exceptional number of novel leads was anticipated using sophisticated large-scale techniques, which would result in a substantial increase in novel entities launched to the market per year. However, this process had some drawbacks which included lack of validation and further optimization of these compounds into actual leads and preclinical candidates.

Such situations fuel the consideration and development of alternative techniques. These techniques have allowed the computational strategies to establish a strong hold in all aspects of drug development process (Jorgensen, 2004; Kitchen, Decornez, Furr, & Bajorath, 2004; Bajorath, 2002; Langer, & Hoffmann, 2001; Walters, Stahl, & Murcko, 1998), such as, virtual screening techniques (Gohlke, & Klebe, 2002) for hit identification and methods for lead optimization. Virtual screening (VS) is a direct and rational drug discovery approach and has the advantage of low cost and effective screening (Moitessier, Englebienne, Lee, Lawandi, & Corbeil, 2008; Shoichet, McGovern, Wei, & Irwin, 2002; Bailey, & Brown, 2001). The expression 'virtual screening' was coined in the late 1990s; however, the techniques involved are much older (Lahana, 1999). In contrast to high-throughput screening (HTS), which is largely driven by phenomenon and technology, in VS, compounds are selected by predicting their binding to a macromolecular target using computer programs. The term 'target' or 'receptor' is used to describe the macromolecule to which a drug binds, which is usually a protein but can also be DNA or RNA. The small molecules or "ligands" may bind to the targets of biological interest. These ligands which are considered for study do not necessarily exist and their 'testing' does not include consumption of valuable resources. Thus, experimental hindrances like low solubility, etc. that can interfere with experimentally applied assay conditions are not needed to be considered for the initial computational screening (Klebe, 2006). VS require a prerequisite knowledge of the spatial arrangement and energetic conditions which are responsible for the binding of a particular ligand molecule to the receptor under investigation. Therefore, either the three-dimensional (3D) structure of the target or, at the very least, a rigid reference ligand with a known bioactive conformation mapping out the putative receptor binding site must be available (Langer, & Krovat, 2003). This defines VS as a knowledge-driven approach. VS can be classified into ligand-based and structure-based methods. When a set of active ligand molecules is known but very little or no information of the structure is available, then the ligand-based methods are employed. This includes pharmacophore modelling (Yu, et al., 2015; Sindhu, & Srinivasan, 2015; Vuorinen, et al., 2014) and quantitative structure activity relationship (QSAR) methods (Rajkhowa, & Deka, 2014; Rajkhowa, Hussain, Hazarika, Sarmah, & Deka, 2013). When the three-dimensional (3D) structure or model of the target is available, then the structure-based virtual screening (SBVS) methods (Istyastono, et al., 2015; Muralidharan, et al. 2015; Lionta, Spyrou, Vassilatis, & Cournia, 2014) and employed. This process is known as the molecular docking process and is used to place computer-generated representations of a small molecule into a target structure in a variety of positions, conformations and orientations (Kroemer, 2007).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Molecular docking is a method which predicts the preferred orientation of a ligand when bound in an active site to form a stable complex. It is the most common method used as a structure-based drug design. Molecular docking is considered as a very important tool in computer-aided drug discovery and has secured a solid position in structure-based drug design. Programs based on different algorithms have been developed to perform molecular docking studies, which have made docking an increasingly important tool in pharmaceutical research. Many reviews on docking have been published in the past (Meng, Zhang, Mezei, & Cui, 2011; Kontoyianni, Madhav, Suchanek, & Seibel, 2008; Coupez, & Lewis, 2006; Kitchen, et al., 2004; Brooijmans, & Kuntz, 2003; Halperin, Ma, Wolfson, & Nussinov, 2002) and many comparative studies have been conducted to evaluate the relative performances of the programs (Plewczynski, Lazniewski, Augustyniak, & Ginalski, 2010; Li, Li, Cheng, Liu, & Wang, 2010; ten Brink, & Exner, 2009; Cross, et al. 2009).

The molecular docking approach can be used to study the interaction of a small molecule with a protein at the atomic level, which helps in characterizing the behaviour of the small molecule in the binding site of the protein as well as to understand the fundamental biochemical processes (McConkey, Sobolev, & Edelman, 2002). The docking process basically involves two steps: (a) prediction of the conformation of a ligand, its position and orientation within the sites, and (b) assessment of the binding affinity. These two steps are related to sampling methods and scoring schemes, respectively

In many cases, the binding site is known before the docking of the ligands, which significantly increases the docking efficiency. Information about the binding site may also be gathered by comparison of the target protein with protein families having similar function or with proteins which are co-crystallized with other ligands. Many cavity detection programs and online servers are available which can be utilized to identify probable active sites within the proteins. Some of the programs that are available are GRID (Kastenholz, Pastor, Cruciani, Haaksma, & Fox, 2000; Goodford, 1985), POCKET (Levitt, & Banaszak, 1992), SURFNET (Glaser, Morris, Najmanovich, Laskowski, & Thornton, 2006; Laskowski, 1995), PASS (Brady, & Stouten, 2000) and MMC (Mezei, 2003). A docking performed without any assumption about the binding site is known as blind docking.

The lock- and -key theory is considered as the first theory which explains the ligand-receptor binding mechanism. It was proposed by Emil Fischer in 1890 (Fischer, 1894). This theory proposes that the ligand fits into the active site of a receptor just like a key fits into a lock (Figure 1). Biological 'locks' have unique stereochemical features that are necessary for their functioning. Some of the earliest reported docking methods (Kuntz, Blaney, Oatley, Langridge, & Ferrin, 1982) were based on this theory where both the ligand and the receptor were treated as rigid bodies.

In 1958, Daniel Koshland introduced the 'induced-fit theory (Koshland, 1963). This theory proposed that the active site of the protein is continually reshaped by interactions with the ligands as the ligands interact with the protein. Here, both the ligand and target mutually adapt to each other through small conformational changes until an optimal fit is achieved as shown in Figure 2. It suggests that both the ligand and the target should be considered as flexible during docking.

Recently, it has been observed that proteins pre-exist as ensemble of conformational states (Boehr, Nussinov, & Wright, 2009). They can undergo much larger conformational changes in addition to the small induced-fit adaptations (Figure 3). The plasticity of the protein allows it to switch from one state to another.

The algorithms which are available to assess and justify protein ligand interactions is very large. They provide a wide range of techniques to tackle modern structure-based drug design. Despite the large size of this field, we have attempted to summarise and classify the types of docking methods. Types of

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 1. Lock- and -key model of Emil Fischer

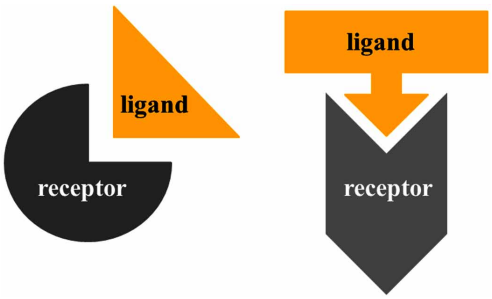


Figure 2. Induced-fit model of Daniel Koshland

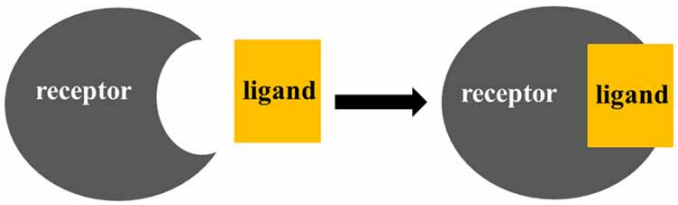
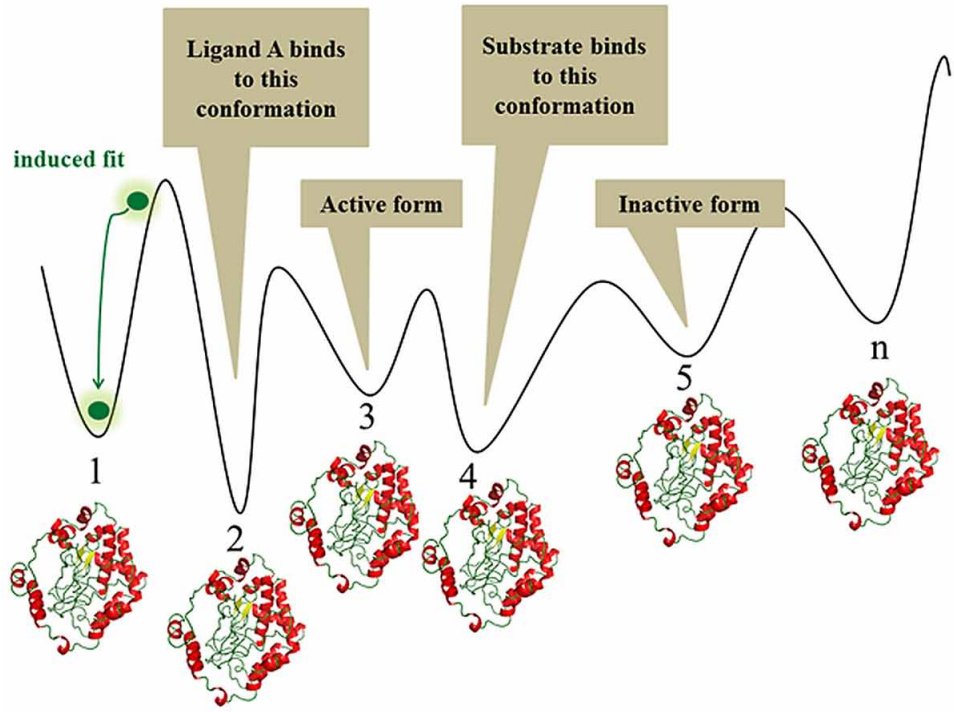


Figure 3. The conformation ensemble model showing the different conformational states of a protein



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

docking methodologies can be divided into: (a) Rigid ligand and rigid receptor docking, (b) Flexible ligand and rigid receptor docking, and (c) Flexible ligand and flexible receptor docking. This chapter will, therefore, highlight the types of docking methods and some of its application in drug discovery.

2. THEORY OF MOLECULAR DOCKING

Molecular docking is a method which predicts the preferred orientation of one molecule (ligand) to a second (protein) when bound to each other to form a stable complex. Docking can be achieved through two interrelated steps: first by sampling conformations of the ligand in the active site of the protein; then ranking these conformations via a scoring function. The sampling algorithms reproduce the experimental binding mode and the scoring functions rank the highest among all generated conformations. From these two viewpoints, we give a brief overview of the basic docking theory.

a. Sampling Algorithms

A huge number of binding modes are possible between a ligand and a protein when their six degrees of translational and rotational freedom as well as conformational degrees of freedom are considered. As a result, it would be computationally too expensive to generate all the possible conformations. To overcome this problem, various sampling algorithms have been developed and widely used in molecular docking softwares (Table 1).

i. Matching Algorithms

Matching algorithms (MA) (Norel, Fischer, Wolfson, & Nussinov, 1994; Fischer, Norel, Wolfson, & Nussinov, 1993; Brint, & Willett, 1987) are based on the molecular shapes. They use the shape features and chemical information of a ligand to map it into an active site of a protein. Here, the protein and ligand are represented as pharmacophores and the distance of these pharmacophores within the protein and the ligand is calculated for a match wherein new ligand conformations are governed by the distance matrix between the pharmacophore and the corresponding ligand atoms. Some chemical properties such as hydrogen-bond donors, hydrogen-bond acceptors and aromatic rings can be taken into account during

Table 1. Some sampling algorithms discussed in the chapter

Algorithms	Characteristic	Ref.
Matching algorithms	Geometry-based, suitable to VS and database enrichment for its high speed	(Norel, et al., 1994; Fischer, et al., 1993; Brint, & Willett, 1987)
Incremental construction	Fragment-based and docking incrementally	(Rarey, et al., 1996; Kuntz, & Leach, 1992; DesJarlais, et al., 1986)
LUDI	Fragment-based methods for the *de novo* design	(Bohm, 1992)
Monte Carlo and Genetic algorithms	Stochastic search	(Morris, et al., 1998; Jones, et al., 1997; Oshiro, et al., 1995; Goodsell, et al., 1993; Hart, & Read, 1992)
Molecular dynamics	For further refinement after docking	(Cornell, et al., 1995 ; Weine, et al., 1984)

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

the match. Matching algorithms for ligand docking are available in programs such as DOCK (Knegtel, Kuntz, & Oshiro, 1997), FLOG (Miller, Kearsley, Underwood, & Sheridan, 1994), LibDock (Diller, & Merz, 2001) and SANDOCK (Burkhard, Taylor, & Walkinshaw, 1998).

ii. Incremental Construction

Incremental construction (IC) (Rarey, Kramer, Lengauer, & Klebe, 1996; Kuntz, & Leach, 1992; Des-Jarlais, Sheridan, Dixon, Kuntz, & Venkataraghavan, 1986) methods are available in programs such as DOCK 4.0 (Ewing, Makino, Skillman, & Kuntz, 2001), FlexX (Kramer, Rarey, & Lengauer, 1999), Hammerhead (Welch, Ruppert, & Jain, 1996), SLIDE (Schnecke, & Kuhn, 2000) and eHiTS (Zsoldos, Reid, Simon, Sadjad, & Johnson, 2006). This method puts the ligand into an active site in a fragmental and incremental fashion. The rotatable bonds are broken so that ligand is divided into several fragments. One of these fragments is selected to dock into the active site first. This anchor is usually the largest fragment or the piece which may have significant functional role or interaction with the protein. The remaining fragments are added incrementally. The flexibility of the ligand is realized by generating different orientations to fit in the active site.

iii. LUDI

LUDI (Bohm, 1992) is a fragment-based method for the *de novo* design and modifications of ligands that may enhance their binding to the target protein. It focuses on the hydrogen bonds and hydrophobic interactions which are formed between the ligand and the protein. This program works a three stage protocol-by calculating the interaction sites of the target protein and ligand that are discrete positions in space, which can either, form hydrogen bonds or fill hydrophobic pockets. The next step involves the fitting of molecular fragments onto the interaction sites. The final stage of the protocol involves the joining of fragments using a database of bridge fragments and the previous fitting algorithm.

iv. Monte Carlo and Genetic Algorithm

Monte Carlo (MC) (Goodsell, Lauble, Stout, & Olson, 1993; Hart, & Read, 1992) and *Genetic Algorithm* (GA) (Morris, et al., 1998; Jones, Willett, Glen, Leach, & Taylor, 1997; Oshiro, Kuntz, & Dixon, 1995) belong to the class of stochastic methods. They search the conformational space by modifying a ligand conformation or a population of ligands randomly. The MC method has been applied to earlier versions of AutoDock (Goodsell, & Olson, 1990), ICM (Abagyan, Totrov, & Kuznetsov, 1994), QXP (McMartin, & Bohacek, 1997) and Affinity (Accelrys Inc., USA.), where it generates the poses of ligands through bond rotation, rigid-body translation or rotation. In order to test the conformations generated by the transformation, an energy-based selection criterion is used. These conformations, when passes the criterion, are then saved and modified to generate the next conformations. This procedure is repeated until the pre-defined quantity of conformations is collected. An advantage of MC technique is that a simple energy function can be used. This energy function does not require any derivative information. On the other hand, genetic algorithm (GA) encodes the degrees of freedom of the ligand as genes or binary strings. The collection of genes, or chromosome, represents the pose of the ligand. Mutation and crossover are two kinds of genetic operators in GA. Mutation makes random changes to the genes,

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

whereas, crossover exchanges genes between two chromosomes. GA has been used in programs such as AutoDock (Goodsell, & Olson, 1990), GOLD (Verdonk, Cole, Hartshorn, Murray, & Taylor, 2003), DIVALI (Clark, & Ajay, 1995) and DARWIN (Taylor, & Burnett, 2000).

v. Molecular Dynamics

Molecular dynamics (MD) (Cornell, et al., 1995; Weine, et al., 1984) involves the calculation of solutions to Newton's equations of motions. A number of programs are available to perform molecular dynamics simulations such as AMBER (Kollman, 1996), GROMACS (van der Spoel, et al., 2005; Berendsen, van der Spoel, & van Drunen, 1995) and CHARMM (Brooks, et al., 1983). From docking point of view, MD simulation represents the flexibility of both the ligand and the protein considering the movement of each atom separately in the field of the rest atoms. However, the movement proceeds in very small steps and thus, have difficulties in stepping over high energy conformational barriers. The quality of the results from a standard MD simulation is extremely dependent on the starting conformation of the system.

b. Scoring Functions

Scoring functions distinguish between the current poses from incorrect ones or good binding ligands from inactive compounds in a reasonable computation time. The scoring function should be able to distinguish the experimental binding modes from all other modes explored through the searching algorithm. Since, a rigorous scoring function will be computationally very expensive, so often the function's complexity is reduced with some loss of accuracy. Scoring functions can be divided into force-field-based, empirical and knowledge-based scoring functions.

The docking programs that are available nowadays utilise the scoring functions in either of the two ways. In the first approach, it uses the full scoring function to rank a protein-ligand conformation. The system is then modified by the search algorithm, and the same scoring function is again applied to rank the new structure. In the second approach, it uses a two stage scoring function. Here, a reduced function is used to direct the search strategy and a more rigorous scoring function is then used to rank the resulting structures (Taylor, Jewsbury, & Essex, 2002). A serious problem faced by scoring functions is that of affinity prediction partly because of the limited treatment of solvation effect. One of the ways to solve this problem is physics-based scoring, which is involved in rescoring or lead optimization to improve the accuracy of binding affinity prediction.

3. TYPES OF DOCKING METHODOLOGIES

a. Rigid Ligand and Rigid Receptor Docking

When the ligand and receptor are both considered as rigid bodies, with the search space being very narrow and considering only the three translational and three rotational degrees of freedom, it is known as rigid docking. Here the protein is treated as rigid and the ligand flexibility is attained by using a pre-computed set of ligand conformations, or by allowing the overlap of the atoms between the protein and the ligand. Programs like HEX (Frishman, & Argos, 1995), and early versions of DOCK (Fischer, et

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

al., 1993), FLOG (Miller, et al., 1994) and some protein-protein docking programs, such as FTDOCK (Gabb, Jackson, & Sternberg, 1997) has adopted this method which keep the ligand and receptor rigid during docking.

HEX is an interactive molecular graphics program for calculating and displaying feasible docking modes of pairs of protein and DNA molecules. Hex calculates the protein-ligand docking by assuming that the ligand and protein are rigid. Here, 3D expansions of real orthogonal spherical polar basis functions are used to model each molecule. This is done in order to encode both surface shape and electrostatic charge and potential distributions. This allows each property to be represented by a vector of coefficients. Hex represents the surface shapes of proteins using a two-term surface skin plus van der Waals steric density model, whereas the electrostatic model is derived from classical electrostatic theory. An overall docking score can be obtained as a function of the six degrees of freedom in a rigid body docking search by writing expressions for the overlap of pairs of parametric functions. This docking score can be interpreted as an interaction energy using suitable scaling factors.

DOCK is the first automated procedure for docking a molecule into a receptor site. DOCK characterizes the ligand and receptor as sets of spheres which could be overlaid my means of a clique detection procedure (Bron, & Kerbosch, 1973). DOCK uses geometrical and chemical matching algorithms as sampling algorithms and the scoring is done by considering steric fit, chemical complementation or pharmacophore similarity. Incremental construction method and exhaustive search are added to the improved versions in order to consider the ligand flexibility.

FLOG generates ligand conformations on the basis of distance geometry. The sets of distances are calculated using a clique-finding algorithm. In FLOG, conformations are scored with a function considering van der Waals, electrostatics, hydrogen bonding and hydrophobic interactions.

b. Flexible Ligand and Rigid Receptor Docking

When the ligand is kept flexible whereas the receptor is considered as rigid, the system follows the flexible ligand and rigid receptor docking. This system follows the induced-fit paradigm where the flexibilities of ligand and receptor are considered. However, keeping the receptor flexible would make the calculation very expensive. Thus, the common method, also a balance between accuracy and computational time, treats the ligand as flexible while receptor is kept rigid during docking. Almost all the docking programs available nowadays have adopted this method, such as AutoDock (Goodsell, & Olson, 1990), Molegro Virtual Docker (Thomsen, & Christensen, 2006) and FlexX (Kramer, et al., 1999).

Molegro Virtual Docker (MVD) is an integrated environment for studying and predicting how ligands interact with macromolecules. The identification of ligand binding modes is done by iteratively evaluating a number of ligand conformations and estimating the energy of their interactions with the target. The highest scoring solutions are returned for further analysis. This program considers the flexibility of the ligand and allows partial flexibility of the receptor (important amino acids present in the binding cavity), considering the cost of computation. MVD uses MolDock as a scoring function. The MolDock scoring function is based on a new hybrid search algorithm, called glided differential evolution. This glided differential evolution technique uses the tripos force-field and combines the differential evolution optimization technique with a cavity prediction algorithm (Storn, & Price, 1995). It is based on a piecewise linear potential (PLP) (Yang, & Chen, 2004).

AutoDock 3.0 (Morris, et al., 1998) uses a genetic algorithm as a global optimiser combined with energy minimisation as a local search method. AutoDock considers the ligand as flexible and the receptor

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

as rigid and represents it as a grid. The scoring function used by AutoDock is based on AMBER forcefield which includes the van der Waals, hydrogen bonding, electrostatic interactions, conformational entropy and desolvation terms. These five terms are weighted using an empirical scaling factor determined using linear regression analysis from a set experimental data. AutoDock 4.0 (Morris, et al., 2009), however, is able to model receptor flexibility by allowing side-chains to move. As a result, interaction of protein-protein docking can be evaluated using this version of AutoDock. AutoDock Vina has been found to significantly improve the average accuracy of the binding mode predictions as compared to AutoDock 4.

FlexX uses an incremental construction algorithm and is one of the most popular programs to perform fragment docking. The base fragment is first docked into the active site by matching hydrogen bond pairs and metal and aromatic ring interactions between the ligand and the protein. The remaining components are then incrementally added according to the predefined rotatable torsion angles which results in the flexibility of the ligand. The scoring function of FlexX is based on Böhm's work (Bohm, 1994). The current version of FlexX includes placing explicit waters into the binding site during the docking procedure (Rarey, Kramer, & Lengauer, 1999) using pre-computed water positions. Some other terms such as electrostatic interactions, directional hydrogen bonds, rotational entropy, and aromatic and lipophilic interactions are also added in the latest extensions. The interactions between functional groups are also taken into account through assigning the type and geometry for groups. Furthermore, receptor flexibility has been included using discrete alternative protein conformations (Claussen, Buning, Rarey, & Lengauer, 2001).

c. Flexible Ligand and Flexible Receptor Docking

Receptor (protein) flexibility plays a very important role in receptor–ligand binding processes and should be considered in computational drug design. However, incorporating the receptor flexibility is a significant challenge in the field of docking. Several approaches have been reviewed (Teodoro, & Kavraki, 2003) in order to incorporate receptor flexibility in ligand docking studies where they have proposed a classification of methods that can be divided into five categories: (a) 'soft' receptors that limit penalties due to steric clashes, (b) selection of a few critical degrees of freedom in the binding site, c) use of multiple receptor structures, d) use of modified molecular simulation methods and e) use of collective degrees of freedom as a new basis of representation for protein flexibility (Totrov, & Abagyan, 2008). A number of methodologies have been proposed to treat the conformational changes in the structure-based docking procedure, which have a significant impact on the accuracy of docking and scoring and they have been extensively reviewed (Kokh, Wade, & Wenzel, 2011).

Molecular dynamics simulations are most commonly used to model all the degrees of freedom in the ligand-receptor complex. However, due to the problem of inadequate sampling and high computational expense, MD simulations cannot be used in the screening of large chemical database. The intrinsic mobility of proteins is related to the ligand binding behaviour (Teague, 2003). Several theoretical models such as conformer selection and conformational induction have been proposed to illustrate the flexible ligand-protein binding process. When a ligand selectively binds to a favourable conformation from a number of protein conformations, it is referred to as conformer selection (Teague, 2003), whereas conformational induction refers to a process in which the ligand converts the protein into a conformation that it would not spontaneously adopt in its unbound state. Various methods are currently available to implement the receptor flexibility (Table 2). They can be classified as:

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 2. Some basic methods for including receptor flexibility along with their advantages and disadvantages

Method	Description	Advantage	Disadvantage	Program
Soft potential	Change vdW to allow for overlap between receptor and ligand atoms	Computational efficiency. Easy to implement and used in combination with other metheds.	Inadequate flexibility. Describe flexibility in an implicit, rude and non-quantitative way.	AutoDock (Goodsell, & Olson, 1990)
Rotamer library	Search side chain library to obtain possible conformations	Relative computational efficiency. Avoid minimization barriers.	Strong dependence on the database used. No backbone flexibility.	ICM (Abagyan, et al., 1994)
Receptor side chain flexibility	Sample both side chain and ligand conformations simultaneously using GA	Relative computational efficiency. Model the effect that ligand make on binding site residues.	Only selected side chains are involved. No backbone flexibility.	GOLD (Verdonk, et al., 2003) AutoDock 4.0 (Morris, et al., 2009)
Ensemble of protein conformations	Docking ligand to a series of receptor structures which represent different conformational states.	Include full and explicit flexibility.	Expensive computational cost. Limited by protein conformations used in sampling.	DOCK (Knegtel, et al., 1997) FlexE (Claussen, et al., 2001)

i. Soft Receptors

The use of soft receptors is the simplest solution to represent some degree of receptor flexibility. The receptor is made softer by reducing the van der Waals contributions to the total energy which allows a larger ligand to fit in a binding site determined experimentally for a small molecule. The reason behind this approach is that the structure of the receptor has some inherent flexibility which allows it to adapt to slightly differently shaped ligands by resorting to small variations in the orientation of binding site chains and backbone positions. If the change in the receptor conformation is small enough, it is assumed that the receptor is capable of such a conformational change, given its large number of degrees of freedom, even though the conformational change itself is not modeled explicitly. The main advantage of using soft receptors is the ease of implementation (docking algorithms stay unchanged) and speed (the cost of evaluating the scoring function is the same as for the rigid case). Examples of soft docking are the LJ 8-4 potential in GOLD and smooth potential in AutoDock 3.0 belong to this class.

ii. Selection of Specific Degrees of Freedom

In order to reduce the complexity of modeling the very large dimensional space representing the full flexibility of the protein, it is possible to obtain an approximate solution by selecting only a few degrees of freedom to model explicitly. The degrees of freedom chosen usually correspond to rotations around single bonds. Rotations around bonds lead to deviations from ideal geometry that resulted in a small energy penalty when compared to deviations from ideality in bond lengths and bond angles. Such rotomer libraries (Leach, 1994; Desmet, Maeyer, Hazes, & Lasters, 1992) include a set of side-chain conformations which are usually determined from statistical analysis of structural experimental data. The main advantage of using rotomer libraries is the relative speed in sampling, and avoiding of minimization

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

barriers. Some examples of programs using rotomer libraries are ICM (Internal Coordinates Mechanics) (Abagyan, et al., 1994) coupled with MC search of the ligand conformation and GOLD (Verdonk, et al., 2003) coupled with GA to dock a flexible ligand to a semi-flexible protein.

iii. Multiple Receptor Structures

One possible way to represent a flexible receptor for drug design applications is the use of multiple static receptor structures. This concept is supported by the currently accepted model that proteins in solution do not exist in a single minimum energy static conformation but are in fact constantly jumping between low energy conformational substrates (Kitao, Hayward, & Go, 1998; Frauenfelder, Sligar, & Wolynes, 1991; Noguti, & Go, 1989). In this way the best description for a protein structure is that of a conformational ensemble (Rich, Bursavich, & Estiarte, 2002) of slightly different protein structures coexisting in a low energy region of the potential energy surface. Here, the ligand is docked separately into a set of rigid protein conformations instead of a single one, and the results are merged depending on the method of choice (Cavasotto, & Abagyan, 2004). One of the main advantages of using multiple structures instead of using a selection of degrees of freedom to represent protein flexibility is that the flexible region is not limited to a specific small region of the protein. Multiple structures allow the consideration of the full flexibility of the protein without the exponential blow up in terms of computational cost that would derive from including all the degrees of freedom of the protein.

FlexE (Claussen, Buning, Rarey, & Lengauer, 2001) represents a flexible binding site by combining multiple receptor structures. The algorithm starts by superimposing the set of conformations available for a given receptor and merging similar parts of the structures. The best combination of substructures is selected with respect to the scoring function based on a comparison between the ligand and each alternative.

iv. Molecular Simulations

To simulate the binding process with as much detail as possible and avoid some of the limitations of previous flexibility models one can use force field based atomistic simulation methods such as MC or MD. MD applies the laws of classical mechanics to compute the motion of the particles in a molecular system, whereas, MC methods are so called because they are based on a random sampling of the conformational space. The main advantage of MC or MD flexibility representations in docking studies is that they are very accurate and can model explicitly all degrees of freedom of the system including the solvent if necessary (Teodoro, & Kavraki, 2003). MC based methods takes into account receptor flexibility by the use of a flexible enzyme binding site whose side chains are submitted to random perturbations.

QXP (Quick Explore) (McMartin, & Bohacek, 1997) is part of the Flo+ program and contains two conceptually different docking algorithms, namely, MCDock and ZipDock. The MCDock algorithm goes through repeated cycles of MC followed by energy minimization in order to generate and refine an ensemble of low-energy ligand poses. The new poses which have low energy and which are dissimilar to the poses already present are added to the ensemble. With the increase in the number of cycles, the size of the perturbations is reduced. This makes the MCDock procedure very efficient in finding low-energy solutions.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

v. Collective Degrees of Freedom

An alternative representation for protein flexibility is the use of collective degrees of freedom which can be determined using different methods such as calculation of normal mode of the receptor (Levy, & Karplus, 1979). The collective degrees of freedom can also be calculated using dimensional reduction method such as principal component analysis (PCA). The receptor flexibility can also be achieved by using the concept of molecular hinges (Sandak, Nussinov, & Wolfson, 1995). These hinges are articulation points placed at specific locations in the protein which allow for relative movement of domains or substructural parts. These hinges are the articulations that are allowed to rotate in three dimensions, implicitly representing rotations about consecutive or nearby bonds. The advantage of collective degrees of freedom is that the protein flexibility is not limited to a specific small region of the protein as was the case when using only selected degrees of freedom. Also, since only a few independent degrees of freedom are used in the optimization procedure, the computational cost is similar to using only selected degrees of freedom and is much less than the cost of techniques that considers all degrees of freedom, such as traditional MD or MC.

Glide (Friesner, et al., 2004) is uses hybrid method to model receptor flexibility by designing a series of hierarchical filters to search the possible poses and orientations of the ligand within the binding site of the receptor. An exhaustive search of the ligand torsion angle space handles the ligand flexibility and the initial ligand conformations are selected based on torsion energies, which are then docked into receptor binding sites with soft potentials. To model the receptor flexibility, a rotamer exploration is used (Sherman, Day, Jacobson, Friesner, & Farid, 2006).

4. APPLICATION EXAMPLES OF MOLECULAR DOCKING IN MODERN DRUG DISCOVERY

Molecular docking has been one of the most widely used techniques in the field of drug design. Although the main application of docking lies in structure-based virtual screening for identification of new compounds towards a particular target protein (Kubinyi, 2006), it is not a stand-alone technique. It should be embedded in a workflow of different *in silico* as well as experimental techniques (Kroemer, 2007). Molecular docking can be combined with other computational techniques and experimental data to analyse the drug metabolism in drug discovery. A few successful applications of docking are discussed in the following.

In recent years, drug delivery by nanomaterials has gained considerable potential application due to its unique quasi one-dimensional structure and electronic properties (Reich,Thomsen, & Maultzsch, 2004). Single-wall carbon nanotubes have been selectively employed to deliver biologically active molecules across the cell membranes. In this context, Deka and co-workers (Saikia, Rajkhowa, & Deka, 2013) have carried out combined density functional theory (DFT) calculations and molecular docking studies on pyrazinamide (PZA) loaded functionalized single-wall carbon nanotube (fSWCNT) with wild type *Mycobacterium tuberculosis* pncA enzyme. They studied the interaction of pyrazinamide (PZA) with fSWCNT as function of nanotube chirality and length using two different approaches of covalent functionalization. They have investigated using DFT formalism the structural stability and electronic properties of prestine zigzag (*n,*0) SWCNT (n = 9, 10, 11, and 12) and functionalized nanotubes with PZA anti-tubercular drug molecules. Their studies revealed that the nanotubes with narrow diameter

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

were thermodynamically more favourable during the course of drug delivery. The decrease in binding energies of PZA/(n,0) SWCNT with increasing diameter indicated that the curvature of nanotube sidewall played a major role in determining the reactivity. Moreover, they also found that highest occupied molecular orbitals (HOMO) contributed from the edge carbon atoms; except for (9,0) SWCNT, wherein it was delocalized along the nanotube sidewall. Their studies further showed that organic moieties like carboxylic and other functional groups could be attached at the reactive ends of the nanotube to enhance the solubility and dispersability of the nanotube in the human body. A detailed analysis on the contribution of HOMO, lowest unoccupied molecular orbitals (LUMO), HOMO – n and LUMO + n have been done which provided valuable insight and significant implications towards the orbital distribution in pristine and functionalized SWCNT systems. The docking studies performed by them suggested that SWCNTs facilitated the effective and target specific delivery of PZA drug within the active site of pncA and that narrow diameter nanotube are well docked rather than the larger diameter tubes (Figure 4). It was also found, that the length of PEG chain is reasonably adequate for the drug delivery of PZA drug within the specific active binding site of pncA protein. Furthermore, it was found that encapsulation facilitated in multiple loading and delivery of PZA molecules within the protein and for all the studied nanosystems conjugated with PZA molecule SWCNT did not bring any conformational change within the protein.

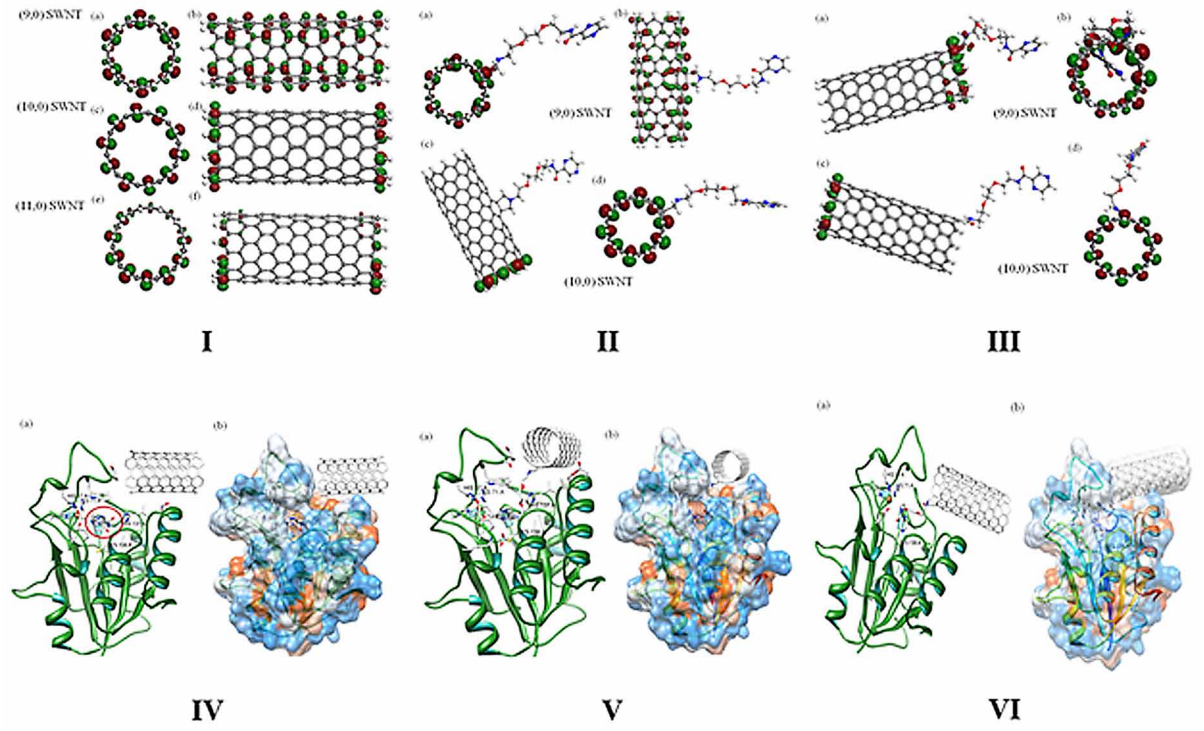
DNA is an important target for the treatment of multiple pathologies such as cancer. In particular, DNA intercalators have often been used as anticancer drugs. However, only a few computational studies have been performed on DNA-intercalator complexes. Very recently, Gilad and Senderowitz (Gilad, & Senderowitz, 2014) have analysed ligand binding site preferences in 63 high resolution DNA-intercalator complexes available in the PDB and found that ligands binds preferentially between G and C and between the C and A base pairs (70% and 11%, respectively). They have also examined the ability of the docking program AUTODOCK to reproduce DNA-ligand complexes. They were able to distinguish between the two major binding modes, namely intercalation and groove binding modes. Using the docking results and a set of ligand-derived molecular descriptors, they were able to develop decision trees to predict the likelihood of successful docking (Figure 5).

High-throughput screening (HTS) and computer-based screening using molecular docking such as VS has been rarely used together on the same target. In this context, Doman *et. al.*(Doman, et al., 2002) have applied both HTS and VS to screen the potential inhibitors of the protein tyrosine phosphatase-1B. For HTS, they have considered a library of approximately 400,000 compounds from a corporate collection. After screening these compounds, they found that only 85 of them have IC_{50} value of less than 100 μM, corresponding to a hit rate of 0.021%. For VS, they performed a docking calculation of 235,000 commercially available molecules using DOCK 3.5. After docking, 1000 top-scoring molecules were considered for further evaluation of which 889 molecules were actually available. After visual inspection, 365 molecules were chosen for testing, of which only 127 molecules were found to be active with IC_{50} less than 100 μM, corresponding to a hit rate of 34.8%. Their work showed that structure-based docking enriched the hit rate by 1700-fold over random screening. Further, they found that the hits from VS and HTS were very different from each other, which implies that combination of VS and HTS may be more helpful for lead discovery.

In another study by Minini *et.al.* (Minini, Álvarez, González, Cerecettob, & Merlino, 2015), the binding modes of 1,2,4-thiadiazol, phenazine and 1,2,6-thiadiazine derivatives to *Tc*TIM have been investigated using molecular docking studies combined with molecular dynamics (MD) simulations. Their results indicated that two of the derivatives showed dimer-disrupting properties towards *Tc*TIM and had allosteric effects in the conformation of the active site. They also showed that one derivative caused

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 4. (I) The front and side view depiction of HOMO plots in pristine (9,0) SWCNT (a, b), pristine (10,0) SWCNT (c, d), pristine (11,0) SWCNT (e, f) each comprising of 5 unit cells, (II) The front and side view depiction of HOMO plots in pristine (9,0) SWCNT (a, b), pristine (10,0) SWCNT (c, d), pristine (11,0) SWCNT (e, f) each comprising of 5 unit cells, (III) The HOMO isosurface of (a) side view and (b) front view depiction of edge functionalized PZA/(9,0) SWCNT, (c) side view and (d) front view depiction of edge functionalized PZA/(10,0) SWCNT, (IV) (a) docked conformation of two PZA drug molecules encapsulated within (6,6) SWCNT comprising of 6 unit cells of SWCNT within the active binding site of pncA protein, (b) the hydrophobic surface of two PZA/(6,6) SWCNT docked within pncA protein., (V) (a) The docked conformation of sidewall functionalized PZA/(9,0) SWCNT comprising of 3 unit cells of SWCNT within the active binding site of pncA protein, (b) hydrophobic surface of PZA/(9,0) SWCNT docked within pncA protein, and (VI) (a) docked conformation of edge functionalized PZA/(9,0) SWCNT comprising of 5 unit cells of SWCNT within the active binding site of pncA protein, (b) the hydrophobic surface of PZA/(9,0) SWCNT docked within pncA protein
Source: Saikia, Rajkhowa, & Deka, 2013.

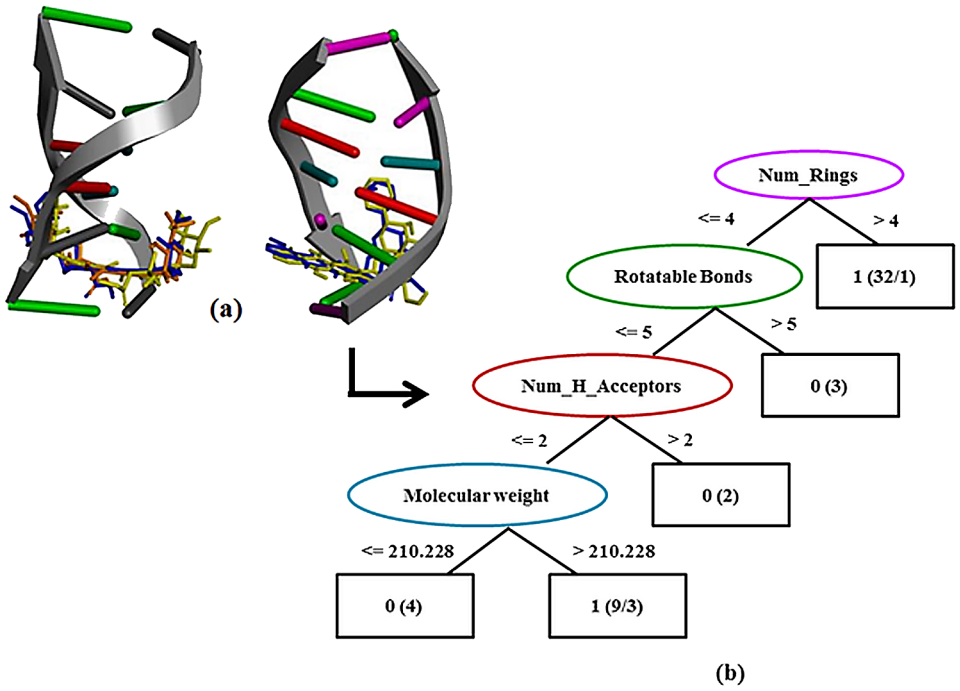


significant decrease in enzyme mobility in both monomers and loss of conformational flexibility. This suggested that this inhibitor could prevent essential motions of the enzyme required for optimal activity.

Rajkhowa et.al. (Rajkhowa, Jha, & Deka, 2015), in their study have used a combination of different computational techniques to virtually screen out four potential drug candidates from a list of possible ones. They employed molecular docking studies to screen molecules from 62 compounds. From this screening they found 32 compounds having good binding modes and top re-rank scores. A comparative study on the chemical properties have been employed using DFT calculations. DFT based global reactivity descriptors (GRD), such as hardness, chemical potential, chemical softness, electrophilicity index, Fukui

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 5. (a) Docked conformations of DNA - ligand complexes. (b) Decision trees for predicting DNA
Source: Gilad, & Senderowitz, 2014.



function, and local philicity calculated at the optimized geometries were used to investigate the usefulness of these descriptors for understanding the reactive nature and sites of the molecules. QSAR equations were built using these descriptors considering these 32 compounds. Four common compounds showing the best correlation and the best docking scores were considered for the ADMET property calculations and their dynamical movements have been studied using molecular dynamics simulations. Their results showed that these four compounds were chemically more active than triclosan and have the potential to inhibit the *Mycobacterium tuberculosis* enoyl acyl carrier protein reductase. Their results showed that combined studies of DFT based QSAR along with docking and MD simulations may provide a better insight towards drug discovery with a higher success rate.

5. CONCLUSION

An extensive summary of the types of docking methods has been presented in this chapter. The receptor flexibility and movement of several key secondary elements of the receptor involving ligand binding and the catalysts, is still a major problem in docking studies. Some of the methods to deal with side chain flexibility have been proven to be effective and adequate in certain areas. An ensemble of proteins is considered as a popular solution for conformer selection with respect to global flexibility. It requires an efficient way to obtain and select reliable protein structures used for docking. Therefore, those structures in which the ligand can fit in should be included in the ensemble. In this regard, Local Move Monte Carlo (LMMC) can be used as an appropriate method for sampling a ligand within loop-containing ac-

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

tive sites. This is because the loop tends to be more flexible and hard to model using existing approaches (Meng, et al., 2011).

Many docking programs exist, but no single program has yet been emerged that has outperformed all others in all cases. Generally, the programs do an adequate job of searching conformational space and generate correct ligand poses (binding modes), however, the scoring functions need improvement. This is evidenced by the fact that the correlation between calculated and observed binding affinities is often low and that separate scoring functions are frequently needed for pose evaluation and affinity prediction. Scoring function is a very important component in docking and therefore, further improvement is needed. Accurate and low computational cost scoring functions may bring docking application to a new stage.

Docking is very useful in reducing a collection of virtual compounds down to a manageable number to be synthesized and in selecting compounds from an external collection. VS is faster than HTS and inexpensive by any standard. It is often recommended to use docking studies in parallel with other techniques to dock to multiple conformations of the target, and to use the docking results to select as many compounds as possible for experimental confirmation.

Considering the known successful applications, the progress made in this field, and the on-going developments, it can be comprehended that the importance and impact of docking will continue to increase significantly.

ACKNOWLEDGMENT

SR acknowledges Department of Science and Technology (DST), New Delhi for providing the DST-INSPIRE Fellowship (Fellow Code: IF110258).

REFERENCES

Abagyan, R., Totrov, M., & Kuznetsov, D. (1994). ICM-A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *Journal of Computational Chemistry*, *15*(5), 488–506. doi:10.1002/jcc.540150503

Bailey, D., & Brown, D. (2001). High-throughput chemistry and structure-based design: Survival of the smartest. *Drug Discovery Today*, *6*(2), 57–59. doi:10.1016/S1359-6446(00)01596-8 PMID:11166243

Bajorath, J. (2002). Integration of virtual and high-throughput screening. *Nature Reviews. Drug Discovery*, *1*(11), 882–894. doi:10.1038/nrd941 PMID:12415248

Berendsen, H. J. C., van der Spoel, D., & van Drunen, R. (1995). GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, *91*(1-3), 43–56. doi:10.1016/0010-4655(95)00042-E

Boehr, D. D., Nussinov, R., & Wright, P. E. (2009). The role of dynamic conformational ensembles in biomolecular recognition. *Nature Chemical Biology*, *5*(11), 789–796. doi:10.1038/nchembio.232 PMID:19841628

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Bohm, H. J. (1992). LUDI: Rule-based automatic design of new substituents for enzyme inhibitor leads. *Journal of Computer-Aided Molecular Design*, *6*, 593–606. doi:10.1007/BF00126217 PMID:1291628

Bohm, H. J. (1994). The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *Journal of Computer-Aided Molecular Design*, *8*, 243–256. doi:10.1007/BF00126743 PMID:7964925

Brady, G. P. Jr, & Stouten, P. F. (2000). Fast prediction and visualization of protein binding pockets with PASS. *Journal of Computer-Aided Molecular Design*, *14*(4), 383–401. doi:10.1023/A:1008124202956 PMID:10815774

Brint, A. T., & Willett, P. (1987). Algorithms for the Identification of Three-Dimensional Maximal Common Substructures. *Journal of Chemical Information and Computer Sciences*, *27*(4), 152–158. doi:10.1021/ci00056a002

Bron, C., & Kerbosch, J. (1973). Algorithm 457: Finding All Cliques of an Undirected Graph. *Communications of the ACM*, *16*(9), 575–576. doi:10.1145/362342.362367

Brooijmans, N., & Kuntz, I. D. (2003). Molecular recognition and docking algorithms. *Annual Review of Biophysics and Biomolecular Structure*, *32*(1), 335–373. doi:10.1146/annurev.biophys.32.110601.142532 PMID:12574069

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, *4*(2), 187–217. doi:10.1002/jcc.540040211

Burkhard, P., Taylor, P., & Walkinshaw, M. D. (1998). An example of a protein ligand found by database mining: Description of the docking method and its verification by a 2.3 A X-ray structure of a thrombin-ligand complex. *Journal of Molecular Biology*, *277*(2), 449–466. doi:10.1006/jmbi.1997.1608 PMID:9514757

Cavasotto, C. N., & Abagyan, R. A. (2004). Protein flexibility in ligand docking and virtual screening to protein kinases. *Journal of Molecular Biology*, *337*(1), 209–225. doi:10.1016/j.jmb.2004.01.003 PMID:15001363

Clark, K. P., & Ajay,. (1995). Flexible ligand docking without parameter adjustment across four ligandreceptor complexes. *Journal of Computational Chemistry*, *16*(10), 1210–1226. doi:10.1002/jcc.540161004

Claussen, H., Buning, C., Rarey, M., & Lengauer, T. (2001). FlexE: Efficient molecular docking considering protein structure variations. *Journal of Molecular Biology*, *308*(2), 377–395. doi:10.1006/jmbi.2001.4551 PMID:11327774

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., & Kollman, P. A. et al. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, *117*(19), 5179–5197. doi:10.1021/ja00124a002

Coupez, B., & Lewis, R. A. (2006). Docking and scoring--theoretically easy, practically impossible? *Current Medicinal Chemistry*, *13*(25), 2995–3003. doi:10.2174/092986706778521797 PMID:17073642

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Cross, J. B., Thompson, D. C., Rai, B. K., Baber, J. C., Fan, K. Y., Hu, Y., & Humblet, C. (2009). Comparison of several molecular docking programs: Pose prediction and virtual screening accuracy. *Journal of Chemical Information and Modeling*, *49*(6), 1455–1474. doi:10.1021/ci900056c PMID:19476350

DesJarlais, R. L., Sheridan, R. P., Dixon, J. S., Kuntz, I. D., & Venkataraghavan, R. (1986). Docking flexible ligands to macromolecular receptors by molecular shape. *Journal of Medicinal Chemistry*, *29*(11), 2149–2153. doi:10.1021/jm00161a004 PMID:3783576

Desmet, J., Maeyer, M. D., Hazes, B., & Lasters, I. (1992). The dead-end elimination theorem and its use in protein sidechain positioning. *Nature*, *356*(6369), 539–542. doi:10.1038/356539a0 PMID:21488406

Diller, D. J., & Merz, K. M. Jr. (2001). High throughput docking for library design and library prioritization. *Proteins*, *43*(2), 113–124. doi:10.1002/1097-0134(20010501)43:2<113::AID-PROT1023>3.0.CO;2-T PMID:11276081

Doman, T. N., McGovern, S. L., Witherbee, B. J., Kasten, T. P., Kurumbail, R., Stallings, W. C., & Shoichet, B. K. et al. (2002). Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. *Journal of Medicinal Chemistry*, *45*(11), 2213–2221. doi:10.1021/jm010548w PMID:12014959

Ewing, T. J., Makino, S., Skillman, A. G., & Kuntz, I. D. (2001). DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *Journal of Computer-Aided Molecular Design*, *15*(5), 411–428. doi:10.1023/A:1011115820450 PMID:11394736

Fischer, D., Norel, R., Wolfson, H., & Nussinov, R. (1993). Surface motifs by a computer vision technique: Searches, detection, and implications for protein-ligand recognition. *Proteins*, *16*(3), 278–292. doi:10.1002/prot.340160306 PMID:8394000

Fischer, E. (1894). Einfluss der configuration auf die wirkungderenzyme. *Berichte der Deutschen Chemischen Gesellschaft*, *27*(3), 2985–2993. doi:10.1002/cber.18940270364

Frauenfelder, H., Sligar, S. G., & Wolynes, P. G. (1991). The energy landscapes and motions of proteins. *Science*, *254*(5038), 1598–1603. doi:10.1126/science.1749933 PMID:1749933

Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., & Shenkin, P. S. et al. (2004). Glide: A new approach for rapid, accurate docking and scoring. 1 Method and assessment of docking accuracy. *Journal of Medicinal Chemistry*, *47*(7), 1739–1749. doi:10.1021/jm0306430 PMID:15027865

Frishman, D., & Argos, P. (1995). Knowledge-based protein secondary structure assignment. *Proteins*, *23*(4), 566–579. doi:10.1002/prot.340230412 PMID:8749853

Gabb, H. A., Jackson, R. M., & Sternberg, M. J. (1997). Modelling protein docking using shape complementarity, electrostatics and biochemical information. *Journal of Molecular Biology*, *272*(1), 106–120. doi:10.1006/jmbi.1997.1203 PMID:9299341

Gilad, Y., & Senderowitz, H. (2014). Docking Studies on DNA Intercalators. *Journal of Chemical Information and Modeling*, *54*(1), 96–107. doi:10.1021/ci400352t PMID:24303988

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Glaser, F., Morris, R. J., Najmanovich, R. J., Laskowski, R. A., & Thornton, J. M. (2006). A method for localizing ligand binding pockets in protein structures. *Proteins*, *62*(2), 479–488. doi:10.1002/prot.20769 PMID:16304646

Gohlke, H., & Klebe, G. (2002). Approaches to the description and prediction of the binding affinity of smallmolecule ligands to macromolecular receptors. *Angewandte Chemie International Edition in English*, *41*(15), 2644–2676. doi:10.1002/1521-3773(20020802)41:15<2644::AID-ANIE2644>3.0.CO;2-O PMID:12203463

Goodford, P. J. (1985). A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *Journal of Medicinal Chemistry*, *28*(7), 849–857. doi:10.1021/jm00145a002 PMID:3892003

Goodsell, D. S., Lauble, H., Stout, C. D., & Olson, A. J. (1993). Automated docking in crystallography: Analysis of the substrates of aconitase. *Proteins*, *17*(1), 1–10. doi:10.1002/prot.340170104 PMID:8234239

Goodsell, D. S., & Olson, A. J. (1990). Automated docking of substrates to proteins by simulated annealing. *Proteins*, *8*(3), 195–202. doi:10.1002/prot.340080302 PMID:2281083

Halperin, I., Ma, B., Wolfson, H., & Nussinov, R. (2002). Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins, 47*, 09–443.

Hart, T. N., & Read, R. J. (1992). A multiple-start Monte Carlo docking method. *Proteins*, *13*(3), 206–222. doi:10.1002/prot.340130304 PMID:1603810

Istyastono, E. P., Kooistra, A. J., Vischer, H. F., Kuijer, M., Roumen, L., Nijmeijer, S., & de Graaf, C. et al. (2015). Structure-based virtual screening for fragment-like ligands of the G protein-coupled histamine H4 receptor. *MedChemComm*, *6*(6), 1003–1017. doi:10.1039/C5MD00022J

Jones, G., Willett, P., Glen, R. C., Leach, A. R., & Taylor, R. (1997). Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology*, *267*(3), 727–748. doi:10.1006/jmbi.1996.0897 PMID:9126849

Jorgensen, W. L. (2004). The many roles of computation in drug discovery. *Science*, *303*(5665), 1813–1818. doi:10.1126/science.1096361 PMID:15031495

Kastenholz, M. A., Pastor, M., Cruciani, G., Haaksma, E. E., & Fox, T. (2000). GRID/CPCA: A new computational tool to design selective ligands. *Journal of Medicinal Chemistry*, *43*(16), 3033–3044. doi:10.1021/jm000934y PMID:10956211

Kitao, A., Hayward, S., & Go, N. (1998). Energy landscape of a native protein: Jumping-among-minima model. *Proteins*, *33*(4), 496–517. doi:10.1002/(SICI)1097-0134(19981201)33:4<496::AID-PROT4>3.0.CO;2-1 PMID:9849935

Kitchen, D. B., Decornez, H., Furr, J. R., & Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nature Reviews. Drug Discovery*, *3*(11), 935–949. doi:10.1038/nrd1549 PMID:15520816

Klebe, G. (2006). Virtual ligand screening: Strategies, perspectives and limitations. *Drug Discovery Today*, *11*(13-14), 580–594. doi:10.1016/j.drudis.2006.05.012 PMID:16793526

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Knegtel, R. M., Kuntz, I. D., & Oshiro, C. M. (1997). Molecular docking to ensembles of protein structures. *Journal of Molecular Biology*, *266*(2), 424–440. doi:10.1006/jmbi.1996.0776 PMID:9047373

Kokh, D. B., Wade, R. C., & Wenzel, W. (2011). Receptor flexibility in small-molecule docking calculations. *WIREs Computational Molecular Science*, *1*(2), 298–314. doi:10.1002/wcms.29

Kollman, P. A. (1996). AMBER 5.0. San Francisco, CA: University of California.

Kontoyianni, M., Madhav, P., Suchanek, E., & Seibel, W. (2008). Theoretical and practical considerations in virtual screening: A beaten field? *Current Medicinal Chemistry*, *15*(2), 107–116. doi:10.2174/092986708783330566 PMID:18220766

Koshland, D. E. Jr. (1963). Correlation of Structure and Function in Enzyme Action. *Science*, *142*(3599), 1533–1541. doi:10.1126/science.142.3599.1533 PMID:14075684

Kramer, B., Rarey, M., & Lengauer, T. (1999). Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. *Proteins*, *37*(2), 228–241. doi:10.1002/(SICI)1097-0134(19991101)37:2<228::AID-PROT8>3.0.CO;2-8 PMID:10584068

Kroemer, R. T. (2007). Structure-based drug design: Docking and scoring. *Current Protein & Peptide Science*, *8*, 312–328. doi:10.2174/138920307781369382 PMID:17696866

Kubinyi, H. (2006). *Computer Applications in Pharmaceutical Research and Development*. New York: John Wiley.

Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R., & Ferrin, T. E. (1982). A geometric approach to macromolecule-ligand interactions. *Journal of Molecular Biology*, *161*(2), 269–288. doi:10.1016/0022-2836(82)90153-X PMID:7154081

Kuntz, I. D., & Leach, A. R. (1992). Conformational analysis of flexible ligands in macromolecular receptor sites. *Journal of Computational Chemistry*, *13*(6), 730–748. doi:10.1002/jcc.540130608

Lahana, R. (1999). How many leads from HTS. *Drug Discovery Today*, *4*(10), 447–448. doi:10.1016/S1359-6446(99)01393-8 PMID:10481138

Langer, T., & Hoffmann, R. D. (2001). Virtual screening: An effective tool for lead structure discovery? *Current Pharmaceutical Design*, *7*(7), 509–527. doi:10.2174/1381612013397861 PMID:11375766

Langer, T., & Krovat, E. M. (2003). Chemical feature-based pharmacophores and virtual library screening for discovery of new leads. *Current Opinion in Drug Discovery & Development*, *6*, 370–376. PMID:12833670

Laskowski, R. A. (1995). SURFNET: A program for visualizing molecular surfaces, cavities, and intermolecular interactions. *Journal of Molecular Graphics*, *13*(5), 323–330. doi:10.1016/0263-7855(95)00073-9 PMID:8603061

Leach, A. R. (1994). Ligand docking to proteins with discrete side-chain flexibility. *Journal of Molecular Biology*, *235*(1), 345–356. doi:10.1016/S0022-2836(05)80038-5 PMID:8289255

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Levitt, D. G., & Banaszak, L. J. (1992). POCKET: A computer graphics method for identifying and displaying protein cavities and their surrounding amino acids. *Journal of Molecular Graphics*, *10*(4), 229–234. doi:10.1016/0263-7855(92)80074-N PMID:1476996

Levy, R. M., & Karplus, M. (1979). Vibrational Approach to the Dynamics of an alpha-Helix. *Biopolymers*, *18*(10), 2465–2495. doi:10.1002/bip.1979.360181008

Li, X., Li, Y., Cheng, T., Liu, Z., & Wang, R. (2010). Evaluation of the performance of four molecular docking programs on a diverse set of protein-ligand complexes. *Journal of Computational Chemistry*, *31*(11), 2109–2125. doi:10.1002/jcc.21498 PMID:20127741

Lionta, E., Spyrou, G., Vassilatis, D. K., & Cournia, Z. (2014). Structure-based virtual screening for drug discovery: Principles, applications and recent advances. *Current Topics in Medicinal Chemistry*, *14*(16), 1923–1938. doi:10.2174/1568026614666140929124445 PMID:25262799

McConkey, B. J., Sobolev, V., & Edelman, M. (2002). The performance of current methods in ligand-protein docking. *Current Science*, *83*, 845–855.

McMartin, C., & Bohacek, R. S. (1997). QXP: Powerful, rapid computer algorithms for structure-based drug design. *Journal of Computer-Aided Molecular Design*, *11*(4), 333–344. doi:10.1023/A:1007907728892 PMID:9334900

Meng, X. Y., Zhang, H. X., Mezei, M., & Cui, M. (2011). Molecular docking: A powerful approach for structure-based drug discovery. *Current Computer-aided Drug Design*, *7*(2), 146–157. doi:10.2174/157340911795677602 PMID:21534921

Mezei, M. (2003). A new method for mapping macromolecular topography. *Journal of Molecular Graphics & Modelling*, *21*(5), 463–472. doi:10.1016/S1093-3263(02)00203-6 PMID:12543141

Miller, M. D., Kearsley, S. K., Underwood, D. J., & Sheridan, R. P. (1994). FLOG: A system to select 'quasi-flexible' ligands complementary to a receptor of known three-dimensional structure. *Journal of Computer-Aided Molecular Design*, *8*(2), 153–174. doi:10.1007/BF00119865 PMID:8064332

Minini, L., Álvarez, G., González, M., Cerecettob, H., & Merlino, A. (2015). Molecular docking and molecular dynamics simulation studies of *Trypanosoma cruzi* triosephosphate isomerase inhibitors. Insights into the inhibition mechanism and selectivity. *Journal of Molecular Graphics & Modelling*, *58*, 40–49. doi:10.1016/j.jmgm.2015.02.002 PMID:25829097

Moitessier, N., Englebienne, P., Lee, D., Lawandi, J., & Corbeil, C. R. (2008). Towards the development of universal, fast and highly accurate docking/scoring methods: A long way to go. *British Journal of Pharmacology*, *153*(S1), S7–S26. doi:10.1038/sj.bjp.0707515 PMID:18037925

Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., & Olson, A. J. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry*, *19*(14), 1639–1662. doi:10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, *30*(16), 2785–2791. doi:10.1002/jcc.21256 PMID:19399780

Muralidharan, A. R., Selvaraj, C., Singh, S. K., Sheu, J. R., Thomas, P. A., & Geraldine, P. (2015). Structure-Based Virtual Screening and Biological Evaluation of a Calpain Inhibitor for Prevention of Selenite-Induced Cataractogenesis in an in Vitro System. *Journal of Chemical Information and Modeling*, *55*(8), 1686–1697. doi:10.1021/acs.jcim.5b00092 PMID:26270943

Noguti, T., & Go, N. (1989). Structural basis of hierarchical multiple substates of a protein. I: Introduction. *Proteins*, *5*(2), 97–103. doi:10.1002/prot.340050203 PMID:2748581

Norel, R., Fischer, D., Wolfson, H. J., & Nussinov, R. (1994). Molecular surface recognition by a computer vision-based technique. *Protein Engineering*, *7*(1), 39–46. doi:10.1093/protein/7.1.39 PMID:8140093

Oshiro, C. M., Kuntz, I. D., & Dixon, J. S. (1995). Flexible ligand docking using a genetic algorithm. *Journal of Computer-Aided Molecular Design*, *9*(2), 113–130. doi:10.1007/BF00124402 PMID:7608743

Patil, S. A. (2012). Role of Medicinal Chemist in the Modern Drug Discovery and Development. *Organic Chemistry: Current Research*, *1*, e110.

Plewczynski, D., Lazniewski, M., Augustyniak, R., & Ginalski, K. (2010). Can we trust docking results? Evaluation of seven commonly used programs on PDB bind database. *Journal of Computational Chemistry*, *32*(4), 742–755. doi:10.1002/jcc.21643 PMID:20812323

Rajkhowa, S., & Deka, R. C. (2014). DFT based QSAR/QSPR models in the development of novel anti-tuberculosis drugs targeting Mycobacterium tuberculosis. *Current Pharmaceutical Design*, *20*(27), 4455–4473. doi:10.2174/13816128196661311181658 24 PMID:24245759

Rajkhowa, S., Hussain, I., Hazarika, K. K., Sarmah, P., & Deka, R. C. (2013). Quantitative structure-activity relationships of the antimalarial agent artemisinin and some of its derivatives - a DFT approach. *Combinatorial Chemistry & High Throughput Screening*, *16*(8), 590–602. doi:10.2174/1386207311316080002 PMID:23597248

Rajkhowa, S., Jha, A. N., & Deka, R. C. (2015). Anti-tubercular drug development: Computational strategies to identify potential compounds. *Journal of Molecular Graphics & Modelling*, *62*, 56–68. doi:10.1016/j.jmgm.2015.09.007 PMID:26386453

Rarey, M., Kramer, B., & Lengauer, T. (1999). The particle concept: Placing discrete water molecules during protein-ligand docking predictions. *Proteins*, *34*(1), 17–28. doi:10.1002/(SICI)1097-0134(19990101)34:1<17::AID-PROT3>3.0.CO;2-1 PMID:10336380

Rarey, M., Kramer, B., Lengauer, T., & Klebe, G. (1996). A fast flexible docking method using an incremental construction algorithm. *Journal of Molecular Biology*, *261*(3), 470–489. doi:10.1006/jmbi.1996.0477 PMID:8780787

Reich, S., Thomsen, C., & Maultzsch, J. (2004). Carbon nanotubes: basic concepts and physical properties. Wiley-VCH.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Rich, D. H., Bursavich, M. G., & Estiarte, M. A. (2002). Discovery of nonpeptide, peptidomimetic peptidase inhibitors that target alternate enzyme active site conformations. *Biopolymers*, *66*(2), 115–125. doi:10.1002/bip.10231 PMID:12325161

Saikia, N., Rajkhowa, S., & Deka, R. C. (2013). Density functional and molecular docking studies towards investigating the role of single-wall carbon nanotubes as nanocarrier for loading and delivering of pyrazinamide antitubercular drug on to pncA protein. *Journal of Computer-Aided Molecular Design*, *27*(3), 57–276. doi:10.1007/s10822-013-9638-6

Sandak, B., Nussinov, R., & Wolfson, H. J. (1995). An automated computer vision and robotics-based technique for 3-D flexible biomolecular docking and matching. *Computer Applications in the Biosciences*, *11*, 87–99. PMID:7796279

Schnecke, V., & Kuhn, L. A. (2000). Virtual Screening with Solvation and Ligand-Induced Complementarity. *Perspectives in Drug Discovery and Design*, *20*(1), 171–190. doi:10.1023/A:1008737207775

Sherman, W., Day, T., Jacobson, M. P., Friesner, R. A., & Farid, R. (2006). Novel procedure for modeling ligand/receptor induced fit effects. *Journal of Medicinal Chemistry*, *49*(2), 534–553. doi:10.1021/jm050540c PMID:16420040

Shoichet, B. K., McGovern, S. L., Wei, B., & Irwin, J. J. (2002) Hits, leads and artifacts from virtual and high throughput screening. In *Proceedings of the International Beilstein Workshop*. Molecular Informatics: Confronting Complexity.

Sindhu, T., & Srinivasan, P. (2015). Identification of potential dual agonists of FXR and TGR5 using e-pharmacophore based virtual screening. *Molecular BioSystems*, *11*(5), 1305–1318. doi:10.1039/C5MB00137D PMID:25787676

Storn, R., & Price, K. (1995). Differential Evolution - A Simple and Efficient Adaptive Scheme for Global Optimization over Continuous Spaces. Technical Report. Berkley, CA: International Computer Science Institute.

Taylor, J. S., & Burnett, R. M. (2000). DARWIN: A program for docking flexible molecules. *Proteins*, *41*(2), 173–191. doi:10.1002/1097-0134(20001101)41:2<173::AID-PROT30>3.0.CO;2-3 PMID:10966571

Taylor, R. D., Jewsbury, P. J., & Essex, J. W. (2002). A review of protein-small molecule docking methods. *Journal of Computer-Aided Molecular Design*, *16*(3), 151–166. doi:10.1023/A:1020155510718 PMID:12363215

Teague, S. J. (2003). Implications of protein flexibility for drug discovery. *Nature Reviews. Drug Discovery*, *2*(7), 527–541. doi:10.1038/nrd1129 PMID:12838268

ten Brink, T., & Exner, T. E. (2009). Influence of protonation, tautomeric, and stereoisomeric states on proteinligand docking results. *Journal of Chemical Information and Modeling*, *49*(6), 1535–1546. doi:10.1021/ci800420z PMID:19453150

Teodoro, M. L., & Kavraki, L. E. (2003). Conformational flexibility models for the receptor in structure based drug design. *Current Pharmaceutical Design*, *9*(20), 1635–1648. doi:10.2174/1381612033454595 PMID:12871062

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Thomsen, R., & Christensen, M. H. (2006). MolDock: A new technique for high-accuracy molecular docking. *Journal of Medicinal Chemistry*, *49*(11), 3315–3321. doi:10.1021/jm051197e PMID:16722650

Totrov, M., & Abagyan, R. (2008). Flexible ligand docking to multiple receptor conformations: A practical alternative. *Current Opinion in Structural Biology*, *18*(2), 178–184. doi:10.1016/j.sbi.2008.01.004 PMID:18302984

van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. C. (2005). GROMACS: Fast, Flexible and Free. *Journal of Computational Chemistry*, *26*(16), 1701–1718. doi:10.1002/jcc.20291 PMID:16211538

Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W., & Taylor, R. D. (2003). Improved protein-ligand docking using GOLD. *Proteins*, *52*(4), 609–623. doi:10.1002/prot.10465 PMID:12910460

Vuorinen, A., Engeli, R., Meyer, A., Bachmann, F., Griesser, U. J., Schuster, D., & Odermatt, A. (2014). Ligand-based pharmacophore modeling and virtual screening for the discovery of novel 17β-hydroxysteroid dehydrogenase 2 inhibitors. *Journal of Medicinal Chemistry*, *57*(14), 5995–6007. doi:10.1021/jm5004914 PMID:24960438

Walters, W. P., Stahl, M. T., & Murcko, M. A. (1998). Virtual screening - an overview. *Drug Discovery Today*, *3*(4), 160–178. doi:10.1016/S1359-6446(97)01163-X

Weine, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., & Weiner, P. et al. (1984). New force field for molecular mechanical simulation of nucleic acids and proteins. *Journal of the American Chemical Society*, *106*(3), 765–784. doi:10.1021/ja00315a051

Welch, W., Ruppert, J., & Jain, A. N. (1996). Hammerhead: Fast, fully automated docking of flexible ligands to protein binding sites. *Chemistry & Biology*, *3*(6), 449–462. doi:10.1016/S1074-5521(96)90093-9 PMID:8807875

Yang, J. M., & Chen, C. C. (2004). GEMDOCK: A generic evolutionary method for molecular docking. *Proteins*, *55*(2), 288–304. doi:10.1002/prot.20035 PMID:15048822

Yu, R., Wang, J., Wang, R., Lin, Y., Hu, Y., Wang, Y., & Lin, Z. et al. (2015). Combined pharmacophore modeling, 3D-QSAR, homology modeling and docking studies on CYP11B1 inhibitors. *Molecules (Basel, Switzerland)*, *20*(1), 1014–1030. doi:10.3390/molecules20011014 PMID:25584832

Zsoldos, Z., Reid, D., Simon, A., Sadjad, B. S., & Johnson, A. P. (2006). eHiTS: An innovative approach to the docking and scoring function problems. *Current Protein & Peptide Science*, *7*(5), 421–435. doi:10.2174/138920306778559412 PMID:17073694

This work was previously published in Methods and Algorithms for Molecular Docking-Based Drug Design and Discovery edited by Siavoush Dastmalchi, Maryam Hamzeh-Mivehroud, and Babak Sokouti, pages 196-219, copyright year 2016 by Medical Information Science Reference (an imprint of IGI Global).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 36
Scoring Functions of Protein-Ligand Interactions

Zhiqiang Yan
Changchun Institute of Applied Chemistry, China

Jin Wang
Stony Brook University, USA

ABSTRACT

Scoring function of protein-ligand interactions is used to recognize the "native" binding pose of a ligand on the protein and to predict the binding affinity, so that the active small molecules can be discriminated from the non-active ones. Scoring function is widely used in computationally molecular docking and structure-based drug discovery. The development and improvement of scoring functions have broad implications in pharmaceutical industry and academic research. During the past three decades, much progress have been made in methodology and accuracy for scoring functions, and many successful cases have be witnessed in virtual database screening. In this chapter, the authors introduced the basic types of scoring functions and their derivations, the commonly-used evaluation methods and benchmarks, as well as the underlying challenges and current solutions. Finally, the authors discussed the promising directions to improve and develop scoring functions for future molecular docking-based drug discovery.

INTRODUCTION

The experiment-based discovery and development of novel drugs are highly complex, time-consuming and expensive (Dickson, 2004). As the number of determined three-dimensional protein structures grows and the power of high-performance computing improves, structure-based drug design and discovery is gaining prominence to identify lead compounds. This field has experienced a rapid growth in academia and pharmaceutical industry. Molecular docking is the most popular computational method employed in structure-based drug design and discovery (Kitchen, 2004; Meng, 2011; Bello, 2013).

Molecular docking mimics the recognition process in which a small molecule (ligand) translates, rotates, and twists thoroughly in the active site of a macromolecule (normally protein), with the goal of finding the most favorable binding mode measured by a scoring function of the binding interactions.

DOI: 10.4018/978-1-5225-0549-5.ch036

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Thus, the protein-ligand docking method can be described as a combination of a search algorithm and a scoring function. The search algorithm generates a number of possible poses that fit the ligand into the binding pocket of the protein. The scoring function scores each pose and then ranks the different poses that are generated by the search algorithm.

In the past three decades, multiple search algorithms were employed in the molecular docking and different types of scoring functions were developed (Dias, 2008; Guo, 2014; Liu, 2015). Various docking programs based on different search algorithms and scoring functions were developed to perform molecular docking studies and applications in structure-based drug design and discovery (Plewczynski, 2011; Sousa, 2013). There have been a mounting number of successful stories with docking-based virtual screening (Tuccinardi, 2009; Villoutreix, 2009; Danishuddin, 2015). Those successes are the outcome of the concerted applications from sampling algorithms, scoring functions as well as filtering criteria. In this chapter, the authors would like to focus on the introduction of scoring functions.

BACKGROUND

Experimentally, the binding affinity (free energy) ($\Delta G_{Bind} = G_{Complex} - G_{Protein} - G_{Ligand}$) of a protein-ligand complex can be given by the determined dissociation constant (K_d) with the equation $\Delta G_{Bind} = RTlnK_d$, where R is the ideal gas constant, and T is the temperature. Computationally, there are a wide variety of different methods capable of predicting the binding affinity of protein-ligand complex. Relatively accurate binding methods, such as free energy perturbation (FEP), thermodynamic integration (TI), linear interaction energy (LIE), Molecular mechanics Poisson-Boltzmann (or Generalized Born) surface area (MM-PB(GB)SA) continuum solvation models (Kollman, 1993; Aqvist, 1994; Gutiérrez-de, 2012; Kollman, 2000), require large computational cost, especially in the computational virtual screening of small molecule databases. Instead, the scoring functions adopt various assumptions and simplifications with approximate mathematical methods to estimate the binding free energy between the protein and the ligand; they are fast in calculating the protein-ligand interactions. Thus, the accuracy is the main task of developers in improving and developing scoring functions.

The accuracy of a scoring function can be evaluated in accordance with its applications in the structure-based drug design and discovery (Huang, 2010; Li, 2014). The first one refers to the ability to rank the computationally generated docking poses, and identify the top ranked one as the "native" binding pose which should be close to experimentally determined structure. The second one refers to the ability to estimate the binding affinities which should be proportional to the experimentally determined binding affinities. The third one refers to the ability to discriminate the true binders against inactive ones in the virtual database screening according to computational binding scores. With the appropriate evaluation methods, these three kinds of abilities can be validated by testing on the protein-ligand benchmarks or with blind assessment exercises.

Scoring functions are either implemented in docking programs or stand-alone in scoring and ranking the generated poses or ligands. So far, the scoring functions have grown into a large diverse family with more than one hundred, whereas, many scoring functions employ common methodologies. Therefore, the scoring functions can be roughly grouped into four main categories: force-field-based, empirical, knowledge-based, and machine-leaning-based scoring functions (Liu, 2015). In addition, to take the advantages and balance the deficiencies of different scoring functions, the consensus scoring technique

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

has been introduced to improve the probability of finding correct solutions by combining the scores from multiple scoring functions.

Though the molecular docking is playing an in increasing role in structure-based drug design and discovery, important challenges are still existed, such as water molecules in binding site, solvation effect, protein flexibility, treatment of entropy, covalent docking, metal ions, and binding specificity (Yan, 2013; Yuriev, 2013; Brás, 2014). Some challenges lie in the side of scoring functions and common to all types of scoring functions. By addressing these obstacles, many attempts have been made to the scoring functions to increase the robustness of reliability and accuracy in their applications, which have paved the way for future directions of improving and developing new scoring functions.

SCORING FUNCTIONS

Classes of Scoring Functions

1. Force-Field-Based Scoring Function

Since the pioneer works by Martin Karplus and coworkers employed molecular mechanics force fields to model biological macromolecules (Karplus, 1976; Brooks, 1983), force-field-based scoring functions have been used for more than two decades to compute the binding affinity between the protein and the ligand (Meng, 1992; Kollman, 1993; Goodsell, 1996; Makino, 1997; Jones, 1997; Kramer, 1999; Yin, 2008). Force-field-based scoring functions generally quantify the sum of two energies, the protein-ligand interaction energy and the internal energy of the ligand. The functional form of the internal ligand energy is very similar to the protein-ligand interaction energy. The energies are normally composed of a van der Waals and an electrostatic energy terms. The general functional form is like

$$S = E_{PL} + E_{L} \tag{1}$$

where E_{PL} represents the energy between the protein and ligand and E_{L} represents the internal energy of the ligand. Different force-field-based scoring functions have similar energy forms with a few different extension energy terms as shown in Table 1. The van der Waals energy term can be given by the Lennard-Jones potential function, the parameters of the Lennard-Jones potential vary by using 12-6 or 8-4 Lennard-Jones potential to adjust the "hardness" of the potential. The electrostatic term is accounted for by a Coulombic formulation with a distance-dependent dielectric function. The effect of solvent is implicitly considered by introducing a simple distance dependent dielectric constant in the Coulombic formulation, which lessens the contribution from charge-charge interactions by mimicking the screening effect. The extensions of general force-field-based scoring functions include a torsional entropy term for ligands and the inclusion of explicit protein-ligand hydrogen-bonding terms. The torsional entropy term is to account for the entropy associated with the rotatable bonds of the ligands. For example, immobilization of a rotatable bond on binding leads to loss of its torsional (or rotational) entropy. The hydrogen-bonding term is modeled to increase the potential of specific ligand binding with noticeably different ways. For example, G-Score (Kramer, 1999) includes different hydrogen-bonding terms de-

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 1. Examples of force-field-based scoring functions

Name	Functional Form
AutoDock4.2 (Morris, 2009)	$S=W_{vdw}\sum_{i,j}\left(\frac{A_{ij}}{r_{ij}^{12}}-\frac{B_{ij}}{r_{ij}^{6}}\right)+W_{hb}\sum_{i,j}E(t)\left(\frac{C_{ij}}{r_{ij}^{12}}-\frac{D_{ij}}{r_{ij}^{10}}\right)+W_{ele}\sum_{i,j}\frac{q_iq_j}{\varepsilon(r_{ij})r_{ij}}+W_{sol}\sum_{i,j}(S_iV_j+S_jV_i)e^{(-r_{ij}^2/2\sigma^2)}$
Dock (Makino, 1997)	$S=\sum_{i,j}\left[\left(\frac{A_{ij}}{r_{ij}^{12}}-\frac{B_{ij}}{r_{ij}^{6}}\right)+332.0\frac{q_iq_j}{\varepsilon(r_{ij})r_{ij}}\right]$
GoldScore (Jones, 1997)	$S=\sum_{i,j}\left[\left(\frac{A_{ij}}{r_{ij}^{a}}-\frac{B_{ij}}{r_{ij}^{b}}\right)+332.0\frac{q_iq_j}{\varepsilon(r_{ij})r_{ij}}\right]+(E_{hb})_{optional}$
D-Score (Meng, 1992)	$S=\sum_{i,j}\left[\left(\frac{A_{ij}}{r_{ij}^{12}}-\frac{B_{ij}}{r_{ij}^{6}}\right)+332.0\frac{q_iq_j}{\varepsilon(r_{ij})r_{ij}}\right]$
G-Score (Jones, 1997)	$S=\sum_{i,j}\left[\left(\frac{A_{ij}}{r_{ij}^{a}}-\frac{B_{ij}}{r_{ij}^{b}}\right)+332.0\frac{q_iq_j}{\varepsilon(r_{ij})r_{ij}}\right]+\sum_{lig}\frac{1}{2}V\left[1+\frac{n}{\lvert n\rvert}\cos(\lvert n\rvert\omega)\right]+E_{hb}$
Atoms i and j are either from the protein and ligand respectively or both from the ligand; A, B, C and D are van der Waals parameters for given atoms, C and D are specialized for hydrogen bonds with oxygen and nitrogen; Deferent W are the weighting constants for the energy terms; E(t) is the angle weighting factor at angel t from ideal hydrogen bonding geometry; ε is the dielectric constant; E_{hb} is the hydrogen bonding term; a and b means that the protein-ligand interaction energy and the internal energy would take different Lenard-Jones parameters; n, w and V are the parameters for the torsional energy in the ligand. For more details, readers are suggested to look into the related citations.	

pending on the nature and geometry of the interaction while AutoDock represents all of the hydrogen bonds by a 12-10 Lennard-Jones potential with an angular weighting factor.

Force-field-based scoring functions were primarily formulated to model enthalpy gas-phase contributions to the energetics. Thus, the major limitations of force-field-based scoring functions are the lack of solvation and entropic terms which are important contributors to the binding free energy between the protein and the ligand. This is in part corrected through the inclusion of additional terms. For example, G-Score includes a torsional entropy term to estimate the conformational entropy loss of ligand upon binding. A straightforward way to account for the solvent effect is to treat water molecules explicitly such as free energy perturbation (FEP) and thermodynamic integration (TI). Alternatively, implicit solvent models include Poisson-Boltzmann (PB) or Generalized Born (GB) continuum solvation models (Kollman, 2000), were developed for scoring use in molecular docking. These methods compute energies as ensemble averages in an attempt to incorporate more physics at the expense of significantly increased computational cost. In practice, these methods are often used in post-scoring of docking programs.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

2. Empirical Scoring Function

The empirical scoring function was first proposed by Böhm in 1992 (Böhm, 1992). An empirical scoring function computes the fitness of protein-ligand binding by summing up the contributions of a number of individual terms. The functional form of empirical scoring functions is often simpler than force-field-based scoring functions, although many of the individual contributing terms have counterparts in the force-field molecular mechanics terms. Generally, the empirical scoring functions decompose the overall binding free energy into several energetic terms as

$$S = \sum_i W_i E_i \tag{2}$$

The coefficients W_i for various terms E_i are obtained from regression analysis using experimentally determined binding free energies and sometimes structural information. Representative examples of empirical scoring functions are given in Table 2. Though the basic expressions are similar between different empirical scoring functions, the energetic terms have rather different implementations. For

Table 2. Examples of empirical scoring functions

Name	Functional Form
LUDI (Böhm, 1992)	$S=\Delta G_{hb}\sum_{hb} f(\Delta R,\Delta\alpha)+\Delta G_{ion}\sum_{ion} f(\Delta R,\Delta\alpha)+\Delta G_{hp}\sum_{hp}\mid A_{hp}\mid +\Delta G_{rot}N_{rot}+\Delta G_0$
ChemScore (Eldridge, 1997)	$S=\Delta G_{hb}\sum_{hb} f(\Delta R,\Delta\alpha)+\Delta G_{met}\sum_{met} f(\Delta R,\Delta\alpha)+$ $\Delta G_{lipo}\sum_{lipo} f(\Delta R)+\Delta G_{rot}\sum_{rot} f(p_{nl},p'_{nl})+\Delta G_0$
F-Score (Rarey, 1996)	$S=\Delta G_{hb}\sum_{hb} f(\Delta R,\Delta\alpha)+\Delta G_{ion}\sum_{ion} f(\Delta R,\Delta\alpha)+\Delta G_{arom}\sum_{arom} f(\Delta R)+\Delta G_{con}\sum_{con} f(\Delta R,\Delta\alpha)$ $+\Delta G_{rot}N_{rot}+\Delta G_0$
GlideScore (Wang, 2002)	$S=C_{\text{lipo-lipo}}\sum f(r_{lr})+C_{hb-neut-neut}\sum g(\Delta r)h(\Delta\alpha)+C_{hb-neut-ch}\sum g(\Delta r)h(\Delta\alpha)$ $+C_{hb-ch-ch}\sum g(\Delta r)h(\Delta\alpha)+C_{\text{max-}met\text{-}ion}\sum f(r_{lr})+C_{rot}H_{rot}+C_{pol-hp}V_{pol-hp}+C_{coul}E_{coul}$ $+C_{vdw}E_{vdw}+E_{sol}$
The scoring above uses energy terms including hydrogen bonding (hb), ionic (ion), hydrophobic (hp), ligand rotational entropy (rotor), contact (con), lipophilic (lipo), metal (met) Coulomb (coul), van der Waals (vdw) and solvation components. The scoring functions differ in energy terms and corresponding regression coefficients. Different ΔG are obtained by multiple linear regressions, ΔG_0 is a regression constant, and the energy term is calculated with a function f or g or h which depends on the angular and/or a distance. GlideScore modifies and extends the ChemScore by classifying hydrogen bonding term into several weighted components depending on whether the donor and the acceptor are neutral (neut) or charged (ch), and incorporating the Coulomb (coul) and van der Waals (vdw) as well as solvation (sol) component. For more details, the readers are suggested to look into the related citations.	

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

instance, LUDI function computes hydrophobic contributions based on the molecular surface area; Glide XP and Fresno (Rognan, 1999) functions include the solvation term to account for the desolvation effect; F-Score function takes into account the interaction between aromatic groups.

Both force-field-based and empirical scoring functions decompose protein-ligand binding free energy into individual energy terms. The major difference between them is that force-field-based scoring function employs the theoretical framework, including the energy function and the associated parameters, from well-established models; whereas the empirical scoring function usually adopts a flexible functional form which is obtained from regression fit. Several force-field-based methods have introduced empirical parameters to reconcile the contributions of its energy terms just like an empirical scoring function. For example, AutoDock4 (Huey, 2007) calibrated the energy terms using a large set of diverse protein-ligand complexes (see Table 1).

The main drawback of the empirical scoring functions is their dependence on the experimental data set used to perform regression analyses and fitting. This means that based on the limited training set, it is rather difficult to establish a comprehensive and consistent description of all possible weighting factors in protein-ligand binding within the framework of an empirical scoring function. i.e. there is no guarantee that an empirical scoring functions is able to predict the binding affinity of the ligands structurally very different from those used in the training set. Nowadays, researchers are able to use several thousands of protein-ligand complexes at a better quality to calibrate or test their scoring functions. The doubt is largely removed while the remaining concern is that the public experimental data are usually measured under different conditions.

3. Knowledge-Based Scoring Function

Since the pioneer work by Tanaka and Scheraga (Tanaka, 1976), a large number of knowledge-based scoring functions have been developed and widely applied to protein structure prediction and protein-ligand studies. The knowledge-based scoring function is also known as statistical potentials. It employs energy potentials that are derived from the structural information embedded in experimentally determined atomic structures (Miyazawa, 1985; Sippl, 1990). The principle behind knowledge-based scoring functions is simple. The interactions of protein-ligand complexes are modeled using pairwise potentials. The pairwise potentials are directly obtained from the occurrence frequency of atom pairs in a database using the inverse Boltzmann relation. The potentials are computed by

$$S = \sum w(r) \tag{3}$$

$$w(r) = -k_B T \ln[g(r)] = -k_B T \ln\left[\frac{f(r)}{f^*(r)}\right] \tag{4}$$

where w(r) is the pairwise potential between protein and ligand, k_B is the Bolzmann constant, T is the absolute temperature of the system, f(r) is the occurrence frequency of the atom pair at distance r and f*(r) is the averaged occurrence frequency in the reference state. If an interatomic distance occurs more often than the average value in the reference state where interatomic interactions are zero, the pairwise

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

potential is assumed as a favorable contact, and vice versa. A challenge in deriving knowledge-based scoring functions is how to achieve the reference state (Thomas, 1996).

The various knowledge-based scoring functions differ between each other in the training sets of protein-ligand complexes used to obtain the potentials, the form of the energy function, the definition of protein and ligand atom types, distance cutoff, and the approximation of the reference state. For example, SMoG uses 13 atom types, two distance intervals and a reference state determined by a self-consistent method (Ishchenko, 2002). SPA uses 22 atom types, 16 distance intervals and a reference state circumvented by Monte Carlo optimization (Yan, 2012). Some methods have combined knowledge-based potentials with solvation and entropy terms (Huang, 2010; Gohlke, 2000; Yan, 2015). In this sense, the boundary between knowledge-based and empirical scoring functions was blurred. The representative knowledge-based scoring functions are SMoG (DeWitte, 1996), PMF(Muegge, 1999), DrugScore (Gohlke, 2000), BLEEP (Mitchell, 1999), ITScore (Huang, 2006), SPA (Yan, 2012).

Unlike force-field-based and empirical scoring functions, knowledge-based potentials attempt to capture all of the energetic factors in protein-ligand interaction implicitly with the pairwise potentials. The major advantage of the knowledge-based scoring functions is their computational simplicity. The deduction of the potentials required for the scoring function only requires the knowledge of a set of protein-ligand complex structures, and such knowledge is relatively rich and still increasing due to the contributions from structural biologists. This overcomes the disadvantage of knowledge-based scoring function that its derivation is based on structural information implicitly encoded in the limited sets of protein-ligand complex structures. In addition, these scoring functions can be easily applied to systems that were not used for the setup of the scoring function, which makes them attractive to apply for very large compound databases.

4. Machine-Learning-Based Scoring Function

Machine-learning-based (ML-based) scoring function is relatively new compared to above three classes of scoring functions since it was first emerged in 2004 (Deng, 2004). These methods employ quantitative structure-activity relationship (QSAR) analysis into protein-ligand interaction evaluation. The machine-learning techniques are applied to derive statistical models that compute protein-ligand binding scores, once the properties of the ligand and the protein as well as their interaction patterns are coded with certain descriptors. The starting point of the method is generally a large pool of descriptors including the descriptors accounting for specific interactions (electrostatic interactions, hydrogen bonds, or aromatic stacking), geometrical descriptors (surface or shape properties), and conventional ligand-based descriptors (molecular weight, number of rotatable bonds, etc.). A variety of machine-learning algorithms, such as neural network, Bayesian classifiers, random forest, and support vector machine, are employed for variable selection. Similar to empirical scoring functions, the method also needs a training set of protein-ligand complexes with known structures and binding data to derive their final models. Typical examples include NNScore (Durrant, 2010), RF-Score (Ballester, 2010), $SFCscore^{RF}$(Zilian, 2013) and ID-Score (Li, 2013).

ML-based scoring functions have attracted attentions and discussions recently due to their good performance in predicting the binding affinity (Ballester, 2010; Kramer, 2010; Ballester, 2011; Zilian, 2013; Ballester, 2014). ML-based scoring functions achieved Pearson correlation coefficients over 0.700 for the testing benchmark, which is superior to other types of scoring functions (shown in Table 4 in the next section). However, Gabel et al. argued that the ML-based scoring functions perform very poorly in

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

the docking test (Gabel, 2014). It means ML-based scoring functions are unable to identify the "native" pose out of the docking decoys. The reason could be that the ML-based functions are over-trained on descriptors that do not depict protein-ligand interactions but interaction-independent atom-pair counts which are insensitive to docking poses. Thus, appropriate improvement is needed to address the issue and extensive validations are necessary for the ML-based scoring functions.

5. Consensus Scoring Function.

Consensus scoring function in essence is not a type of scoring function but combines the scores from multiple scoring functions. It is first proposed by Charifson et al. (1999) and become a new trend in this field. The aim of consensus scoring is to balance errors in single scores and improve the probability of identifying "true" modes/binders. Generally, they are employed as a post-processing step after docking runs. The crucial implementation in consensus scoring is the design of an appropriate consensus scoring strategy of individual scores so that the "true" modes/binders can be discriminated from others (Wang, 2001; Clark, 2002). Commonly used consensus scoring strategies include vote-by-number, number-by-number, rank-by-number, average rank, linear combination, etc (Oda, 2006). Representative examples of consensus scoring are MultiScore (Terp, 2001), X-Cscore (Wang, 2002), GFscore (Betzi, 2006), SCS (Teramoto, 2007a), SeleX-CS (Bar-Haim, 2009).

Many studies have showed that employing consensus scoring approaches can improve the performance by compensating for the deficiencies of multiple individual scoring functions and tend to reduce the number of false positives that are identified by individual scoring functions (Bissantz, 2000; Feher, 2006). It is worth noting that the potential advantage of consensus scoring is limited when the predictions from different scoring functions are highly correlated.

Evaluation of Scoring Functions

1. Evaluation Methods

As introduced, there are three applications for the scoring functions, identifying the "native" pose out of the conformational decoys, predicting the binding affinity of the protein-ligand interactions and virtual database screening. In line with these three applications, the performance of a scoring function can be evaluated with three related but independent methods. Researchers have named them as docking power, scoring power and screening power (Li, 2014).

The docking power requires the structural comparisons of the docking decoys with the "native" conformation of the ligand. It determines the performance of a scoring function in identifying the best-scored decoy as the "native" binding mode. Normally, if RMSD value of the best-scored binding pose to the "native" pose is less than a predefined cutoff, it is considered as a successful identification of the "native" pose by the scoring function. However, this criterion could be problematic in some cases due to its simplicity. For example, small or nearly symmetrical ligands are likely to obtain relative small RMSD values even when they are randomly placed in a small active site. On the contrary, large flexible ligands are likely to obtain the large RMSD values. To overcome these limitations, several alternative methods have been proposed for pose evaluations, such as relative displacement error (RDE) (Abagyan, 1997), interaction-based accuracy classification (IBAC) (Kroemer, 2004), real space R-factor (RSR) (Yusuf, 2008), and generally applicable replacement for RMSD (GARD) (Baber, 2009).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

The scoring power refers to the prediction accuracy of a scoring function on how closely it can estimate the binding affinities in comparison with the experimental measurements. The predicted scores with scaling by scoring functions usually reproduce the relative values of experimental binding affinities. Therefore, the correlations between the predicted binding scores and experimentally measured binding affinities were generally used to evaluate the accuracy of binding affinity prediction. The higher the correlation coefficient is, the better the scoring power is. The commonly-used correlation is the Pearson correlation (CP) which is computed as

$$C_P = \frac{\sum (S_i - < S >)(G_i - < G >)}{\sqrt{\sum (S_i - < S >)^2} \sqrt{\sum (G_i - < G >)^2}} \quad (5)$$

where S_i is the score of a protein-ligand complex computed by a certain scoring function, G_i is the experimental binding affinity. According to the comparative assessment of scoring functions (Li, 2014), the performance of scoring functions in binding affinity is limited with the correlations below 0.70. This could be resulted from two aspects. Firstly, current scoring functions are still in need of improvement as stated in this chapter. Second, the correlation coefficient is dependent on the accuracy and reliability of experimental affinities. Experimental affinities are collected from diverse sources where the binding constants are likely measured under different conditions; these differences can lead to inherent errors among the experimental affinities although carefully cultivated by the researchers (Wang, 2004; Block, 2006; Dunbar Jr, 2011). The inherent errors of the collected experimental binding constant limit the correlation, and make the improvement much harder when correlation coefficient is relatively high.

The screening power determines the ability of a scoring function in discriminating the true binders to a given protein against a large database of compounds. Virtual database screening tests whether or not a scoring function is able to rank the known binders/inhibitors above the inactive compounds in a database. The enrichment test is often taken to quantify the performance of a scoring function in virtual database screening. The enrichment is defined as the accumulated rate of active inhibitors/binders found above a certain percentile of the ranked database that includes the active binders and inactive ligands. A higher enrichment at a fixed percentage of the ranked database can be taken to indicate a better scoring function. Other measurements for virtual database screening are AUC (Area Under the Curve, such as AOC (receiver operating characteristic) curve) and KS (Kolmogorov-Smirno) test, which are more appropriate when the number of inactive ligands is comparable to the number of active binders.

Ideally, an accurate scoring function should be able to perform well above three evaluations. However, most of the existing scoring functions usually perform well on only one or two of the evaluations and fail on others (Li, 2014). In addition, the performances of scoring functions are sometimes sensitive to the testing set. Therefore, to fully evaluate the performance of a scoring function, the above three evaluations are suggested to be carried out with more than one test sets.

2. Evaluation Exercise and Benchmark

An objective and publicly available exercise or benchmark for assessing the performance of the scoring functions can provide practical guidance for the users to make appropriate choices. It can also elucidate the common weakness in the scoring functions for future improvements. The commonly-used exercises

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

and benchmarks in the research community are CSAR (Community Structure-Activity Resource, http://www.csardock.org/) (Smith, 2011) and SAMPL (Statistical Assessment of the Modeling of Proteins and Ligands, http://www.eyesopen.com/SAMPL) (Nicholls, 2008), CASF (Comparative Assessment of Scoring Functions, http://www.pdbbind-cn.org/) (Cheng, 2009) and DUD (Directory of Useful Decoys, http://dud.docking.org/) (Huang, 2006; Irwin, 2008). CSAR and SAMPL exercise results are blind to participants before the exercises were finished. The exercises are normally carried out annually and the benchmarks are often updated and improved with new and higher-quality data. It is worth mentioning that scoring functions are often applied in combination with molecular docking methods to predict ligand binding poses and ligand binding affinities or to identify active compounds through virtual screening. Thus, the evaluation exercises or benchmarks are designed to assess not only the scoring functions but also the docking programs.

3. Comparative Evaluation

By applying the evaluation methods and using the benchmark, researchers have recently compared the performance of different scoring functions testing on the benchmark CASF-2013 (Li, 2014; Yan, 2015). CASF-2013 is composed of 195 protein-ligand complexes with known structures and experimental affinities. To evaluate the scoring functions, a set of binding poses were generated for each protein-ligand complex with RMSDs ranging from 0 to 10 angstroms. The binding poses were scored by the scoring functions including Force-field-based, empirical and knowledge-based scoring functions. The testing results of docking power and scoring power were listed in the Table 3 and Table 4.

It can be seen that all the scoring functions predict the affinities with correlation coefficients lower than 0.7. This suggests that all the testing scoring functions are still not accurate in predicting binding affinities. Compared to the scoring power, more than half scoring functions achieves success rate over 80% in the docking power when top three poses are considered, suggesting that the scoring functions have higher capability in identifying the "native'' conformation. Based on the comparative evaluations of different scoring functions, there is no clear advantage for a particular class of scoring function since the performance of the scoring functions in the same class is diversely ranked in the tables. It is worth noting that SPA-SE ranks best for both docking power and scoring power, as stated in the paper (Yan, 2015), this is because SPA-SE incorporates solvation effect in the scoring and binding specificity in the optimization, which are discussed next.

CHALLENGES AND SOLUTIONS

The molecular docking involves concerted performance of sampling and scoring the binding poses. To this end, there are some major challenges, such as protein flexibility in the sampling, implicit or explicit solvation effect of the binding site, entropy contribution to the binding free energy, covalent bonding between the protein and the ligand, metal ion coordination during binding, binding specificity of molecular recognition, structural water molecules in the binding site and so on. Most of the challenges are related to both sampling and scoring methods and a few are beyond them. Here, the authors introduced the major challenges on the side of the scoring function and recent solutions, such as incorporating the implicit or explicit solvation effect of the binding site, incorporating the entropy contribution to

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 3. Comparative evaluations of scoring functions in the docking power

Scoring Function	Classification	Success Rate (%)		
		Top One	Top Two	Top Three
SPA-SE	Knowledge-based	83.1	89.2	92.8
SPA	Knowledge-based	82.1	88.2	90.8
ChemPLP@GOLD	Empirical	81.0	86.7	89.7
ChemScore@GOLD	Empirical	77.9	83.1	88.2
GlideScore-SP	Empirical	78.5	85.6	87.7
LigScore2@DS	Empirical	76.9	84.1	86.7
PLP1@DS	Empirical	76.9	84.1	86.2
Alpha-HB@MOE	Empirical	75.4	82.6	86.2
GlideScore-XP	Empirical	74.4	82.6	85.6
ASP@GOLD	Knowledge-based	71.8	81.5	87.2
GoldScore@GOLD	Force-field-based	71.3	81.0	85.6
LUDI1@DS	Empirical	59.0	75.4	83.1
Affinity-dG@MOE	Empirical	63.1	74.9	81.0
ChemScore@SYBYL	Empirical	59.5	69.2	75.4
London-dG@MOE	Empirical	59.5	73.8	78.5
X-Score^{HM}	Empirical	61.0	73.3	77.9
G-Score@SYBYL	Force-field-based	45.1	61.5	72.3
Jain@DS	Empirical	48.2	62.1	70.8
PMF@SYBYL	Knowledge-based	51.8	60.0	66.7
PMF04@DS	Knowledge-based	51.8	62.6	66.2
ASE@MOE	Empirical	51.3	60.0	63.6
D-Score@SYBYL	Force-field-based	18.5	29.7	42.6

The results of scoring function SPA-SE and SPA are obtained from the paper (Yan, 2015) and others are from the paper (Li, 2014). For more details, the readers are suggested to look into the citations.

the binding free energy and taking account of binding specificity in the development strategy. Scoring functions accounting for such contributions are more accurate than those without these considerations.

Solvation Effect

It is believed that water molecules actively participate in and affect the process of protein-ligand binding in aqueous solvent (Dill, 2005; Levy, 2006; Berne, 2009; Yan, 2015). The hydration shell that forms around the protein and the ligand is of particular importance to their binding activity. During the binding process, both solvated protein and ligand will be partly dehydrated and the water layers will be gradually excluded into the bulk water, making the ligand diffuse into the active site of the protein in the solvent. The role of water molecules not only assists the formation of the protein-ligand interface but also probably makes the active site more adaptable for ligand binding than other possible binding sites, smoothing the binding energy landscape (Papoian, 2003; Levy, 2006). In this sense, the solvation effect is assumed to be a regulator to the energetics as well as the specificity of ligand binding.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 4. Comparative evaluations of scoring functions in the scoring power

Scoring Function	Classification	C_P	D_S	N_C
SPA-SE	Knowledge-based	0.662	1.69	195
SPA	Knowledge-based	0.656	1.70	195
X-Score^{HM}	Empirical	0.614	1.78	195
ChemScore@SYBYL	Empirical	0.592	1.82	195
ChemPLP@GOLD	Empirical	0.579	1.84	195
PLP1@DS	Empirical	0.568	1.86	195
G-Score@SYBYL	Force-field-based	0.558	1.87	195
ASP@GOLD	Knowledge-based	0.556	1.88	195
ASE@MOE	Empirical	0.544	1.89	195
ChemScore@SYBYL	Empirical	0.536	1.90	189
D-Score@SYBYL	Force-field-based	0.526	1.92	195
Alpha-HB@MOE	Empirical	0.511	1.94	195
LUDI3@DS	Empirical	0.487	1.97	195
GoldScore@GOLD	Force-field-based	0.483	1.97	189
Affinity-dG@MOE	Empirical	0.482	1.98	195
LigScore2@DS	Empirical	0.456	2.02	190
GlideScore-SP	Empirical	0.452	2.03	169
Jain@DS	Empirical	0.408	2.05	191
PMF@DS	Knowledge-based	0.364	2.11	194
GlideScore-XP	Empirical	0.277	2.18	164
London-dG@MOE	Empirical	0.242	2.19	195
PMF@SYBYL	Knowledge-based	0.221	2.20	191

The results of scoring function SPA-SE and SPA are obtained from the paper (Yan, 2015) and others are from the paper (Li, 2014). C_P is the Pearson correlation between experimental and predicted affinities computed by Equation 5; the D_S is the standard deviation in the linear correlation; The N_C is the number of complexes used in the Pearson correlation calculation. For more details, the readers are suggested to look into the citations.

To account for the solvation effect, much progress has been made in the development of both explicit and implicit solvent models (Feig, 2004; Purisima, 2013). Explicit model with inclusion of water molecules can provide detailed structural and energetic information; whereas it significantly increases computational cost (Levy, 1998). Especially, it is unaffordable to screen a large library of small molecules in the docking-based drug discovery. The charge-dependent implicit solvent models, such as PB/GBSA (Poisson-Boltzmann/Generalized Born Surface Area) models, are widely taken as post-processing methods to compute free energies based on molecular dynamics simulations or energy minimizations (Kollman, 2000; Lindstrom, 2011); whereas, these models are rarely incorporated into the scoring function for the process of protein-ligand docking (Zou, 1999). Alternatively, the charge-independent implicit solvent models have the advantage of less computational cost due to simple additive expression since the models are based on solvent accessible surface area (SASA) of each atom and the atomic solvation parameter (Eisenberg, 1986). However, the atomic solvation parameters are usually independently optimized (Wang, 2001; Hou, 2002; Guvench, 2004; Hou, 2005), which are not transferable to the scoring functions of

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

protein-ligand interactions which are described with different classifications of atomic types compared to those used in the solvent models.

Given above difficulties, two knowledge-based scoring functions SPA-SE (Yan, 2015) and ITScore (Huang, 2010) have effectively combined statistical atomic-pair potentials and the atomic solvation energy. The atomic solvation energy is represented by the charge-independent implicit solvent model which accounts for the contribution of solvation energy to the protein-ligand binding based on the change of SASA (Solvent Accessible Surface Area) upon binding. SPA-SE optimized the energetic parameters by maximizing the affinity and specificity predictions of ligand binding, while ITScore employed an iterative method to derive the parameters until the predicted energy converges to the experimentally observed one. Both methods have showed improvements in predicting the binding affinity and identifying the "native" pose.

Entropy Contribution

Conformational entropy is an important component to the free energy change upon binding of a ligand to the protein (Finkelstein, 1989; Murray, 2002; Salaniwal, 2007). The entropy contribution to the binding free energy essentially is arisen from the reduction of the translational and rotational degrees of freedom in the ligand and from changes in the flexibility of the protein during binding, as well as the water molecules involved (Hermans, 1997; Amzel, 1997). The computation of the entropic contribution is very time-consuming (Baron, 2009). Therefore, incorporating such effects in an efficient scoring function is still an important challenge in the field of protein-ligand docking. Some attempts to incorporate entropy in scoring functions have been reported in the literatures which implemented them in the re-scoring schemes (Ruvinsky, 2005; Ruvinsky, 2007; Lee, 2008).

Ruvinsky and Kozintsev (2005) proposed a way to estimate contributions of the translational, rotational and torsional entropy to protein-ligand affinity. In their approach, A number of docking runs are carried out and the generated conformations are clustered by structural similarity. The size of each cluster is then measured to estimate the entropic contribution based on the assumption that large clusters of conformations are indicative of favorable entropic contributions of the local energy landscapes. With this approach, testing scoring functions have shown some improvements in identifying the "native" binding pose (Ruvinsky, 2007). Lee and Seok (2008) introduced a probability function to analyze the populations of diverse binding poses in the context of statistical mechanics. The approach estimates the contribution of the state represented by a sampled conformation of the configurational integral (Xiang, 2002). The approach in combination with 11 scoring functions has resulted in improved accuracy in binding mode prediction. Both approaches are relatively low computational cost and need the combination with other pre-existing scoring functions. However, only the entropy contribution resulted from the ligand were accounted for in these studies. Some other MD-based methods were proposed to estimate the conformational entropy of the protein (Andricioae, 2001; Baron, 2006; Baron, 2008), but they are challenging to be widely employed in the scoring functions due to highly computational demanding.

In addition to the entropic contributions of the ligand and protein, other key components still require our attention. For example, the entropic contribution of the solvent molecules, particularly those at the binding pocket is normally forgotten (Chiba, 2011; Breiten, 2013). The very limited movement of buried waters in the protein-ligand complex and the huge variation arising from desolvation make this an important issue.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Binding Specificity

Biomolecular recognition requires both stability and specificity to realize its function (Janin, 1995; Havranek, 2003; Wang, 2003; Wang, 2007; Kortemme, 2004; Bolon, 2005). The stability of the complex is determined by the binding affinity while the specificity is measured by the discrimination against competitive binding complexes. Most scoring functions of protein-ligand binding, whether force-field based, empirical, knowledge-based scoring functions or other quantitative description of protein-ligand interactions, mainly aim to predict the binding affinities. The strategy of developing these methods seeks to optimize the stability based on the combination of energetics and shape complementarity but without explicit consideration of binding specificity.

Practically, the quantification of binding specificity is a challenge since it requires comparison of all the possible binding complexes. To circumvent the challenge, the concept of intrinsic specificity was proposed to describe the binding specificity (Wang, 2003; Wang, 2007, Yan, 2012). According to the funnel-like energy landscape of protein-ligand binding (Wang, 2003; Wang, 2007; Bryngelson, 1995; Janin, 1996; Rejto, 1996; Miller, 2008; Tsai, 1999; Dominy, 2004; Z. Liu, 2004; Levy, 2004), the intrinsic specificity ratio (ISR) was quantified to measure the binding specificity. With the quantification of the binding specificity, Yan and Wang (Yan, 2012) developed a novel scoring function SPA by optimizing both specificity and affinity. The development strategy for SPA is to simultaneously optimize binding specificity and affinity of ligand binding. SPA was validated by testing the benchmark set and the performance was compared with 16 popular scoring functions. The test results showed that SPA outperformed the compared scoring functions. The work built up a new frame to develop and improve the scoring function of protein-ligand interactions by incorporating the binding specificity.

FUTURE RESEARCH DIRECTIONS

With the efforts in the past three decades, scoring functions of protein-ligand interactions have made much progress and have been used for many successful applications. Currently, there are large data sets of protein-ligand complexes publicly available with known structures and binding affinities to improve the quality. Also, quite a few benchmarks have been created so that different scoring functions can be compared in a more objective manner. This will be undoubtedly further enhancing the improvement of current scoring functions.

Despite the numerous successful applications of molecular docking-based drug discovery, a widely accepted fact is that the major weakness of current computational protein-ligand docking lies not in the sampling methods but in the scoring functions (Kitchen, 2004; Danishuddin, 2015). The benchmark testing suggests that current scoring functions can provide promising results in predicting the correct ligand binding pose, but their performance is relatively weak in predicting protein-ligand binding affinities. In virtual screening, it is still challenging for scoring functions to differentiate low- to medium-affinity binders from non-binders. Thus, considerable efforts are still needed to the development and improvement of the scoring function aiming to rapidly and accurately quantify the interaction between the protein and the ligand (Huang, 2010; Wang, 2013; Liu, 2015). As discussed, a current trend in the field is to focus on the inclusion of solvation effect, entropy contributions as well as binding specificity. Scoring functions accounting for those are more accurate but are also computationally more expensive.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

The fast and accurate prediction of a ligand tightly and specifically binding to a target protein will be the final aim of the scoring functions in the docking-based drug discovery and design.

So far, there are a large number of alternatives of the scoring functions, most of them are generic models derived from the large-scale experimental data of protein-ligand complexes and are presumably applicable to all kinds of target classes. However, it is still difficult to develop a scoring function which can perform equally well across many different protein families (Seifert, 2009; Scharfe, 2012; Cheng, 2012). Binding sites and recognition processes have unique features that ultimately render protein-ligand interactions specific, which in turn might require re-calibrating of scoring schemes on a case-by-case basis. Thus, scoring functions can be improved by tailoring them to a specific target site or to multiple related receptors. For example, the kinase family-specific potential of mean force (kinase-PMF) (Xue, 2010) is kinase-specific scoring function adjusted from the original PMF04 (Muegge, 2006). Target-specific scoring functions have shown improved performance and enhanced screening enrichment in the tests (Teramoto, 2007b; Pham, 2008; Seifert, 2008). It is a promising trend to develop target-specific scoring functions in the future since the improvement of generic scoring functions is often limited with the compromise between the accuracy and computational cost.

CONCLUSION

The discovery and development of novel drugs are highly complex, time-consuming and expensive. Docking-based drug discovery and design is gaining prominence as a method to filter the very large chemical space. Molecular docking-based virtual screening becomes routine in both pharmaceutical companies and academic groups for early-stage drug discovery. Scoring function is at the heart of molecular docking by assisting a docking program to efficiently explore the binding space of a ligand. It is also responsible for evaluating the binding affinity once the binding poses are generated. Therefore, the performance of scoring functions has a significant impact on the docking-based drug discovery. A multitude of scoring functions has been reported in the past three decades and new ones are still emerging. The current trend to improve the performance of scoring functions focusing on the inclusion of additional factors to account for solvation and entropic effects, as well as the strategy to quantify the binding specificity and to develop the target-specific scoring functions. As large data sets of protein-ligand complexes with high quality are available to the public, as well as the computation power increasingly improves, researchers will reach better tradeoff between the accuracy and computational cost for the scoring functions in the coming years. It will undoubtedly produce more intriguing and important applications through molecular docking-based drug design and discovery.

REFERENCES

Abagyan, R. A., & Totrov, M. M. (1997). Contact area difference (CAD): A robust measure to evaluate accuracy of protein models. *Journal of Molecular Biology*, *268*(3), 678–685. PMID:9171291

Amzel, L. M. (1997). Loss of translational entropy in binding, folding, and catalysis. *Proteins:* Structure Function and Genetics, *28*(2), 144–149. PMID:9188731

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Andricioaei, I., & Karplus, M. (2001). On the calculation of entropy from covariance matrices of the atomic fluctuations. *The Journal of Chemical Physics*, *115*(14), 6289–6292.

Aqvist, J., Medina, C., & Samuelsson, J.-E. (1994). A new method for predicting binding affinity in computer-aided drug design. *Protein Engineering*, *7*(3), 385–391. PMID:8177887

Baber, J. C., Thompson, D. C., Cross, J. B., & Humblet, C. (2009). GARD: A generally applicable replacement for rmsd. *Journal of Chemical Information and Modeling*, *49*(8), 1889–1900. PMID:19618919

Ballester, P. J., & Mitchell, J. B. (2010). A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking. *Bioinformatics (Oxford, England)*, *26*(9), 1169–1175. PMID:20236947

Ballester, P. J., & Mitchell, J. B. (2011). Comments on leave-cluster-out cross-validation is appropriate for scoring functions derived from diverse protein data sets: Significance for the validation of scoring functions. *Journal of Chemical Information and Modeling*, *51*(8), 1739–1741. PMID:21591735

Ballester, P. J., Schreyer, A., & Blundell, T. L. (2014). Does a more precise chemical description of protein–ligand complexes lead to more accurate prediction of binding affinity? *Journal of Chemical Information and Modeling*, *54*(3), 944–955. PMID:24528282

Bar-Haim, S., Aharon, A., Ben-Moshe, T., Marantz, Y., & Senderowitz, H. (2009). Selex-CS: A new consensus scoring algorithm for hit discovery and lead optimization. *Journal of Chemical Information and Modeling*, *49*(3), 623–633. PMID:19231809

Baron, R., Hunenberger, P. H., & McCammon, J. A. (2009). Absolute single-molecule entropies from quasi-harmonic analysis of microsecond molecular dynamics: Correction terms and convergence properties. *Journal of Chemical Theory and Computation*, *5*(12), 3150–3160. PMID:20011626

Baron, R., & McCammon, J. A. (2008). (Thermo) dynamic role of receptor flexibility, entropy, and motional correlation in protein–ligand binding. *ChemPhysChem*, *9*(7), 983–988. PMID:18418822

Baron, R., van Gunsteren, W. F., & Hünenberger, P. H. (2006). Estimating the configurational entropy from molecular dynamics simulations: Anharmonicity and correlation corrections to the quasi-harmonic approximation. *Trends in Physical Chemistry*, *11*, 87–122.

Bello, M., Mart'ınez-Archundia, M., & Correa-Basurto, J. (2013). Automated docking for novel drug discovery. *Expert Opinion on Drug Discovery*, *8*(7), 821–834. PMID:23642085

Berne, B. J., Weeks, J. D., & Zhou, R. (2009). Dewetting and hydrophobic interaction in physical and biological systems. *Annual Review of Physical Chemistry*, *60*, 85. PMID:18928403

Betzi, S., Suhre, K., Chétrit, B., Guerlesquin, F., & Morelli, X. (2006). GFScore: A general nonlinear consensus scoring function for high-throughput docking. *Journal of Chemical Information and Modeling*, *46*(4), 1704–1712. PMID:16859302

Bissantz, C., Folkers, G., & Rognan, D. (2000). Protein-based virtual screening of chemical databases. 1. evaluation of different docking/scoring combinations. *Journal of Medicinal Chemistry*, *43*(25), 4759–4767. PMID:11123984

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Block, P., Sotriffer, C. A., Dramburg, I., & Klebe, G. (2006). AffinDB: A freely accessible database of affinities for protein–ligand complexes from the pdb. *Nucleic Acids Research*, *34*(suppl 1), D522–D526. PMID:16381925

Böhm, H.-J. (1992). LUDI: Rule-based automatic design of new substituents for enzyme inhibitor leads. *Journal of Computer-Aided Molecular Design*, *6*(6), 593–606. PMID:1291628

Bolon, D. N., Grant, R. A., Baker, T. A., & Sauer, R. T. (2005). Specificity versus stability in computational protein design. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(36), 12724–12729. PMID:16129838

Brás, N., Cerqueira, N., Sousa, S., Fernandes, P., & Ramos, M. (2014). Protein ligand docking in drug discovery. *Protein Modelling*, 249.

Breiten, B., Lockett, M. R., Sherman, W., Fujita, S., Al-Sayah, M., Lange, H., & Whitesides, G. M. et al. (2013). Water networks contribute to enthalpy/entropy compensation in protein–ligand binding. *Journal of the American Chemical Society*, *135*(41), 15579–15584.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, *4*(2), 187–217.

Bryngelson, J. D., Onuchic, J. N., Socci, N. D., & Wolynes, P. G. (1995). Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins: Structure, Function, and Bioinformatics*, *21*(3), 167–195. PMID:7784423

Charifson, P. S., Corkery, J. J., Murcko, M. A., & Walters, W. P. (1999). Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *Journal of Medicinal Chemistry*, *42*(25), 5100–5109. PMID:10602695

Cheng, T., Li, Q., Zhou, Z., Wang, Y., & Bryant, S. H. (2012). Structure-based virtual screening for drug discovery: A problem-centric review. *The AAPS Journal*, *14*(1), 133–141. PMID:22281989

Cheng, T., Li, X., Li, Y., Liu, Z., & Wang, R. (2009). Comparative assessment of scoring functions on a diverse test set. *Journal of Chemical Information and Modeling*, *49*(4), 1079–1093. PMID:19358517

Chiba, S., Harano, Y., Roth, R., Kinoshita, M., & Sakurai, M. (2012). Evaluation of protein-ligand binding free energy focused on its entropic components. *Journal of Computational Chemistry*, *33*(5), 550–560.

Clark, R. D., Strizhev, A., Leonard, J. M., Blake, J. F., & Matthew, J. B. (2002). Consensus scoring for ligand/protein interactions. *Journal of Molecular Graphics & Modelling*, *20*(4), 281–295. PMID:11858637

Cosconati, S., Forli, S., Perryman, A. L., Harris, R., Goodsell, D. S., & Olson, A. J. (2010). Virtual screening with AutoDock: Theory and practice. *Expert Opinion on Drug Discovery*, *5*(6), 597–607. PMID:21532931

Danishuddin, M., & Khan, A. U. (2015). Structure based virtual screening to discover putative drug candidates: Necessary considerations and successful case studies. *Methods (San Diego, Calif.)*, *71*, 135–145. PMID:25448480

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Deng, W., Breneman, C., & Embrechts, M. J. (2004). Predicting protein-ligand binding affinities using novel geometrical descriptors and machine-learning methods. *Journal of Chemical Information and Computer Sciences*, *44*(2), 699–703. PMID:15032552

DeWitte, R. S., & Shakhnovich, E. I. (1996). Smog: De novo design method based on simple, fast, and accurate free energy estimates. 1. Methodology and supporting evidence. *Journal of the American Chemical Society*, *118*(47), 11733–11744.

Dias, R., de Azevedo, J., & Walter, F. (2008). Molecular docking algorithms. *Current Drug Targets*, *9*(12), 1040–1047. PMID:19128213

Dickson, M., & Gagnon, J. P. (2004). Key factors in the rising cost of new drug discovery and development. *Nature Reviews. Drug Discovery*, *3*(5), 417–429. PMID:15136789

Dill, K. A., Truskett, T. M., Vlachy, V., & Hribar-Lee, B. (2005). Modeling water, the hydrophobic effect, and ion solvation. *Annual Review of Biophysics and Biomolecular Structure*, *34*, 173–199. PMID:15869376

Dominy, B. N., & Shakhnovich, E. I. (2004). Native atom types for knowledge-based potentials: Application to binding energy prediction. *Journal of Medicinal Chemistry*, *47*(18), 4538–4558. PMID:15317465

Dunbar, J. B. Jr, Smith, R. D., Yang, C.-Y., Ung, P. M.-U., Lexa, K. W., Khazanov, N. A., & Carlson, H. A. et al. (2011). CSAR benchmark exercise of 2010: Selection of the protein–ligand complexes. *Journal of Chemical Information and Modeling*, *51*(9), 2036–2046. PMID:21728306

Durrant, J. D., & McCammon, J. A. (2010). NNScore: A neural-network-based scoring function for the characterization of protein- ligand complexes. *Journal of Chemical Information and Modeling*, *50*(10), 1865–1871. PMID:20845954

Eisenberg, D., & Mclachlan, A. (1986). Solvation energy in protein folding and binding. *Nature*, *319*(6050), 199–203. PMID:3945310

Eldridge, M. D., Murray, C. W., Auton, T. R., Paolini, G. V., & Mee, R. P. (1997). Empirical scoring functions: I. the development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *Journal of Computer-Aided Molecular Design*, *11*(5), 425–445. PMID:9385547

Feher, M. (2006). Consensus scoring for protein–ligand interactions. *Drug Discovery Today*, *11*(9), 421–428. PMID:16635804

Feig, M., & Brooks,, C. L. III (2004). Recent advances in the development and application of implicit solvent models in biomolecule simulations. *Current Opinion in Structural Biology*, *14*(2), 217–224. PMID:15093837

Finkelstein, A. V., & Janin, J. l. (1989). The price of lost freedom: Entropy of bimolecular complex formation. *Protein Engineering*, *3*(1), 1–3. PMID:2813338

Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., & Mainz, D. T. et al. others. (2004). Glide: A new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy. *Journal of Medicinal Chemistry*, *47*(7), 1739–1749. PMID:15027865

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., & Mainz, D. T. et al. (2006). Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *Journal of Medicinal Chemistry*, *49*(21), 6177–6196. PMID:17034125

Gabel, J., Desaphy, J., & Rognan, D. (2014). Beware of machine learning-based scoring functionsr on the danger of developing black boxes. *Journal of Chemical Information and Modeling*, *54*(10), 2807–2815. PMID:25207678

Gohlke, H., Hendlich, M., & Klebe, G. (2000). Knowledge-based scoring function to predict protein-ligand interactions. *Journal of Molecular Biology*, *295*(2), 337–356. PMID:10623530

Goodsell, D. S., Morris, G. M., & Olson, A. J. (1996). Automated docking of flexible ligands: Applications of autodock. *Journal of Molecular Recognition*, *9*(1), 1–5. PMID:8723313

Guo, L., Yan, Z., Zheng, X., Hu, L., Yang, Y., & Wang, J. (2014). A comparison of various optimization algorithms of protein–ligand docking programs by fitness accuracy. *Journal of Molecular Modeling*, *20*(7), 1–10. PMID:24935106

Gutiérrez-de Terán, H., & Aqvist, J. (2012). Linear interaction energy: method and applications in drug design. In *Computational drug discovery and design* (pp. 305–323). Springer.

Guvench, O., & Brooks, C. L. (2004). Efficient approximate all-atom solvent accessible surface area method parameterized for folded and denatured protein conformations. *Journal of Computational Chemistry*, *25*(8), 1005–1014. PMID:15067676

Havranek, J. J., & Harbury, P. B. (2003). Automated design of specificity in molecular recognition. *Nature Structural Biology*, *10*(1), 45–52. PMID:12459719

Hermans, J., & Wang, L. (1997). Inclusion of loss of translational and rotational freedom in theoretical estimates of free energies of binding. application to a complex of benzene and mutant t4 lysozyme. *Journal of the American Chemical Society*, *119*(11), 2707–2714.

Hou, T., Qiao, X., Zhang, W., & Xu, X. (2002). Empirical aqueous solvation models based on accessible surface areas with implicit electrostatics. *The Journal of Physical Chemistry B*, *106*(43), 11295–11304.

Hou, T., Zhang, W., Huang, Q., & Xu, X. (2005). An extended aqueous solvation model based on atom-weighted solvent accessible surface areas: SAWSA v2. 0 model. *Journal of Molecular Modeling*, *11*(1), 26–40. PMID:15565273

Huang, N., Shoichet, B. K., & Irwin, J. J. (2006). Benchmarking sets for molecular docking. *Journal of Medicinal Chemistry*, *49*(23), 6789–6801. PMID:17154509

Huang, S.-Y., Grinter, S. Z., & Zou, X. (2010). Scoring functions and their evaluation methods for protein–ligand docking: Recent advances and future directions. *Physical Chemistry Chemical Physics*, *12*(40), 12899–12908. PMID:20730182

Huang, S.-Y., & Zou, X. (2006). An iterative knowledge-based scoring function to predict protein–ligand interactions: I. derivation of interaction potentials. *Journal of Computational Chemistry*, *27*(15), 1866–1875. PMID:16983673

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Huang, S.-Y., & Zou, X. (2010). Inclusion of solvation and entropy in the knowledge-based scoring function for protein-ligand interactions. *Journal of Chemical Information and Modeling*, *50*(2), 262–273. PMID:20088605

Huey, R., Morris, G. M., Olson, A. J., & Goodsell, D. S. (2007). A semiempirical free energy force field with charge-based desolvation. *Journal of Computational Chemistry*, *28*(6), 1145–1152. PMID:17274016

Irwin, J. J. (2008). Community benchmarks for virtual screening. *Journal of Computer-Aided Molecular Design*, *22*(3-4), 193–199. PMID:18273555

Ishchenko, A. V., & Shakhnovich, E. I. (2002). Small molecule growth 2001 (SMoG2001): An improved knowledge-based scoring function for protein-ligand interactions. *Journal of Medicinal Chemistry*, *45*(13), 2770–2780. PMID:12061879

Janin, J. (1995). Principles of protein-protein recognition from structure to thermodynamics. *Biochimie*, *77*(7), 497–505. PMID:8589061

Janin, J. (1996). Quantifying biological specificity: The statistical mechanics of molecular recognition. *Proteins: Structure, Function, and Bioinformatics*, *25*(4), 438–445. PMID:8865339

Jones, G., Willett, P., Glen, R. C., Leach, A. R., & Taylor, R. (1997). Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology*, *267*(3), 727–748. PMID:9126849

Karplus, M., & Weaver, D. L. (1976). Protein-folding dynamics. *Nature*, *260*, 404–406. PMID:1256583

Kitchen, D. B., Decornez, H., Furr, J. R., & Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nature Reviews. Drug Discovery*, *3*(11), 935–949. PMID:15520816

Kollman, P. (1993). Free energy calculations: Applications to chemical and biochemical phenomena. *Chemical Reviews*, *93*(7), 2395–2417.

Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., & Chong, L. et al. others. (2000). Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Accounts of Chemical Research*, *33*(12), 889–897. PMID:11123888

Kortemme, T., Joachimiak, L. A., Bullock, A. N., Schuler, A. D., Stoddard, B. L., & Baker, D. (2004). Computational redesign of protein-protein interaction specificity. *Nature Structural & Molecular Biology*, *11*(4), 371–379. PMID:15034550

Kramer, B., Rarey, M., & Lengauer, T. (1999). Evaluation of the flexx incremental construction algorithm for protein–ligand docking. *Proteins: Structure, Function, and Bioinformatics*, *37*(2), 228–241. PMID:10584068

Kramer, C., & Gedeck, P. (2010). Leave-cluster-out cross-validation is appropriate for scoring functions derived from diverse protein data sets. *Journal of Chemical Information and Modeling*, *50*(11), 1961–1969. PMID:20936880

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Kroemer, R. T., Vulpetti, A., McDonald, J. J., Rohrer, D. C., Trosset, J.-Y., Giordanetto, F., & Stouten, P. F. et al. (2004). Assessment of docking poses: Interactions-based accuracy classification (IBAC) versus crystal structure deviations. *Journal of Chemical Information and Computer Sciences*, *44*(3), 871–881. PMID:15154752

Lee, J., & Seok, C. (2008). A statistical rescoring scheme for protein–ligand docking: Consideration of entropic effect. *Proteins: Structure, Function, and Bioinformatics*, *70*(3), 1074–1083. PMID:18076034

Levy, R. M., & Gallicchio, E. (1998). Computer simulations with explicit solvent: Recent progress in the thermodynamic decomposition of free energies and in modeling electrostatic effects. *Annual Review of Physical Chemistry*, *49*(1), 531–567. PMID:9933909

Levy, Y., & Onuchic, J. N. (2006). Water mediation in protein folding and molecular recognition. *Annual Review of Biophysics and Biomolecular Structure*, *35*, 389–415. PMID:16689642

Levy, Y., Wolynes, P. G., & Onuchic, J. N. (2004). Protein topology determines binding mechanism. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(2), 511–516. PMID:14694192

Li, G.-B., Yang, L.-L., Wang, W.-J., Li, L.-L., & Yang, S.-Y. (2013). Id-score: A new empirical scoring function based on a comprehensive set of descriptors related to protein–ligand interactions. *Journal of Chemical Information and Modeling*, *53*(3), 592–600. PMID:23394072

Li, Y., Han, L., Liu, Z., & Wang, R. (2014). Comparative assessment of scoring functions on an updated benchmark: II. Evaluation methods and general results. *Journal of Chemical Information and Modeling*, *54*, 1717–1736. PMID:24708446

Li, Y., Liu, Z., Li, J., Han, L., Liu, J., Zhao, Z.-X., & Wang, R. (2014). Comparative assessment of scoring functions on an updated benchmark: I. Compilation of the test set. *Journal of Chemical Information and Modeling*, *54*, 1700–1716. PMID:24716849

Lindstrom, A., Edvinsson, L., Johansson, A., Andersson, C. D., Andersson, I. E., Raubacher, F., & Linusson, A. (2011). Postprocessing of docked protein-ligand complexes using implicit solvation models. *Journal of Chemical Information and Modeling*, *51*(2), 267–282. PMID:21309544

Liu, J., & Wang, R. (2015). On classification of current scoring functions. *Journal of Chemical Information and Modeling*, *55*(3), 475–482. PMID:25647463

Liu, Z., Dominy, B. N., & Shakhnovich, E. I. (2004). Structural mining: Self-consistent design on flexible protein-peptide docking and transferable binding affinity potential. *Journal of the American Chemical Society*, *126*(27), 8515–8528. PMID:15238009

Makino, S., & Kuntz, I. D. (1997). Automated flexible ligand docking method and its application for database search. *Journal of Computational Chemistry*, *18*(14), 1812–1825.

Meng, E. C., Shoichet, B. K., & Kuntz, I. D. (1992). Automated docking with grid-based energy evaluation. *Journal of Computational Chemistry*, *13*(4), 505–524.

Meng, X.-Y., Zhang, H.-X., Mezei, M., & Cui, M. (2011). Molecular docking: A powerful approach for structure-based drug discovery. *Current Computer-aided Drug Design*, *7*(2), 146. PMID:21534921

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Miller, D. W., & Dill, K. A. (2008). Ligand binding to proteins: The binding landscape model. *Protein Science*, *6*(10), 2166–2179. PMID:9336839

Mitchell, J. B., Laskowski, R. A., Alex, A., & Thornton, J. M. (1999). Bleeppotential of mean force describing protein–ligand interactions: I. generating potential. *Journal of Computational Chemistry*, *20*(11), 1165–1176.

Miyazawa, S., & Jernigan, R. L. (1985). Estimation of effective interresidue contact energies from protein crystal structures: Quasi-chemical approximation. *Macromolecules*, *18*(3), 534–552.

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and autodocktools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, *30*(16), 2785–2791. PMID:19399780

Muegge, I. (2006). PMF scoring revisited. *Journal of Medicinal Chemistry*, *49*(20), 5895–5902. PMID:17004705

Muegge, I., & Martin, Y. C. (1999). A general and fast scoring function for protein-ligand interactions: A simplified potential approach. *Journal of Medicinal Chemistry*, *42*(5), 791–804. PMID:10072678

Murray, C. W., & Verdonk, M. L. (2002). The consequences of translational and rotational entropy lost by small molecules on binding to proteins. *Journal of Computer-Aided Molecular Design*, *16*(10), 741–753. PMID:12650591

Nicholls, A., Mobley, D. L., Guthrie, J. P., Chodera, J. D., Bayly, C. I., Cooper, M. D., & Pande, V. S. (2008). Predicting small-molecule solvation free energies: An informal blind test for computational chemistry. *Journal of Medicinal Chemistry*, *51*(4), 769–779. PMID:18215013

Oda, A., Tsuchida, K., Takakura, T., Yamaotsu, N., & Hirono, S. (2006). Comparison of consensus scoring strategies for evaluating computational models of protein-ligand complexes. *Journal of Chemical Information and Modeling*, *46*(1), 380–391. PMID:16426072

Papoian, G. A., Ulander, J., & Wolynes, P. G. (2003). Role of water mediated interactions in protein-protein recognition landscapes. *Journal of the American Chemical Society*, *125*(30), 9170–9178. PMID:15369374

Pham, T. A., & Jain, A. N. (2008). Customizing scoring functions for docking. *Journal of Computer-Aided Molecular Design*, *22*(5), 269–286. PMID:18273558

Plewczynski, D., La'zniewski, M., Augustyniak, R., & Ginalski, K. (2011). Can we trust docking results? Evaluation of seven commonly used programs on pdbbind database. *Journal of Computational Chemistry*, *32*(4), 742–755. PMID:20812323

Purisima, E., & Sulea, T. (2013). Solvation models: Theory and validation. *Current Pharmaceutical Design*, *20*(20), 3266–3280. PMID:23947651

Rarey, M., Kramer, B., Lengauer, T., & Klebe, G. (1996). A fast flexible docking method using an incremental construction algorithm. *Journal of Molecular Biology*, *261*(3), 470–489. PMID:8780787

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Rejto, P. A., & Verkhivker, G. M. (1996). Unraveling principles of lead discovery: From unfrustrated energy landscapes to novel molecular anchors. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(17), 8945–8950. PMID:8799133

Rognan, D., Lauemøller, S. L., Holm, A., Buus, S., & Tschinke, V. (1999). Predicting binding affinities of protein ligands from three-dimensional models: Application to peptide binding to class i major histocompatibility proteins. *Journal of Medicinal Chemistry*, *42*(22), 4650–4658. PMID:10579827

Ruvinsky, A. M. (2007). Role of binding entropy in the refinement of protein–ligand docking predictions: Analysis based on the use of 11 scoring functions. *Journal of Computational Chemistry*, *28*(8), 1364–1372. PMID:17342720

Ruvinsky, A. M., & Kozintsev, A. (2005). New and fast statistical-thermodynamic method for computation of protein-ligand binding entropy substantially improves docking accuracy. *Journal of Computational Chemistry*, *26*(11), 1089–1095. PMID:15929088

Salaniwal, S., Manas, E. S., Alvarez, J. C., & Unwalla, R. J. (2007). Critical evaluation of methods to incorporate entropy loss upon binding in high-throughput docking. *Proteins: Structure, Function, and Bioinformatics*, *66*(2), 422–435. PMID:17068803

Scharfe, M., Pippel, M., & Sippl, W. (2012). Development of target-biased scoring functions for protein-ligand docking. *Journal of Cheminformatics*, 4(Suppl 1).

Seifert, M. H. (2008). Optimizing the signal-to-noise ratio of scoring functions for protein-ligand docking. *Journal of Chemical Information and Modeling*, *48*(3), 602–612. PMID:18293951

Seifert, M. H. (2009). Targeted scoring functions for virtual screening. *Drug Discovery Today*, *14*(11), 562–569. PMID:19508918

Shoichet, B. K. (2004). Virtual screening of chemical libraries. *Nature*, *432*(7019), 862–865. PMID:15602552

Sippl, M. J. (1990). Calculation of conformational ensembles from potentials of mena force: An approach to the knowledge-based prediction of local structures in globular proteins. *Journal of Molecular Biology*, *213*(4), 859–883. PMID:2359125

Smith, R. D., Dunbar, J. B. Jr, Ung, P. M.-U., Esposito, E. X., Yang, C.-Y., Wang, S., & Carlson, H. A. (2011). CSAR benchmark exercise of 2010: Combined evaluation across all submitted scoring functions. *Journal of Chemical Information and Modeling*, *51*(9), 2115–2131. PMID:21809884

Sousa, S., Cerqueira, N., Fernandes, P., & Ramos, M. (2010). Virtual screening in drug design and development. *Combinatorial Chemistry & High Throughput Screening*, *13*(5), 442–453. PMID:20236061

Sousa, S., Ribeiro, A., Coimbra, J., Neves, R., Martins, S., Moorthy, N., & Ramos, M. et al. (2013). Protein-ligand docking in the new millennium–a retrospective of 10 years in the field. *Current Medicinal Chemistry*, *20*(18), 2296–2314. PMID:23531220

Tanaka, S., & Scheraga, H. A. (1976). Medium-and long-range interaction parameters between amino acids for predicting three-dimensional structures of proteins. *Macromolecules*, *9*(6), 945–950. PMID:1004017

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Teramoto, R., & Fukunishi, H. (2007a). Supervised consensus scoring for docking and virtual screening. *Journal of Chemical Information and Modeling*, *47*(2), 526–534. PMID:17295466

Teramoto, R., & Fukunishi, H. (2007b). Supervised scoring models with docked ligand conformations for structure-based virtual screening. *Journal of Chemical Information and Modeling*, *47*(5), 1858–1867. PMID:17685604

Terp, G. E., Johansen, B. N., Christensen, I. T., & Jørgensen, F. S. (2001). A new concept for multidimensional selection of ligand conformations (Multiselect) and multidimensional scoring (multiscore) of protein-ligand binding affinities. *Journal of Medicinal Chemistry*, *44*(14), 2333–2343. PMID:11428927

Thomas, P. D., & Dill, K. A. (1996). Statistical potentials extracted from protein structures: How accurate are they? *Journal of Molecular Biology*, *257*(2), 457–469. PMID:8609636

Tsai, C.-J., Kumar, S., Ma, B., & Nussinov, R. (1999). Folding funnels, binding funnels, and protein function. *Protein Science*, *8*(6), 1181–1190. PMID:10386868

Tuccinardi, T. (2009). Docking-based virtual screening: Recent developments. *Combinatorial Chemistry & High Throughput Screening*, *12*(3), 303–314. PMID:19275536

Villoutreix, B. O., Eudes, R., & Miteva, M. A. (2009). Structure-based virtual ligand screening: Recent success stories. *Combinatorial Chemistry & High Throughput Screening*, *12*(10), 1000–1016. PMID:20025565

Wang, J., & Verkhivker, G. M. (2003). Energy landscape theory, funnels, specificity, and optimal criterion of biomolecular binding. *Physical Review Letters*, *90*(18), 188101. PMID:12786043

Wang, J., Wang, W., Huo, S., Lee, M., & Kollman, P. A. (2001). Solvation model based on weighted solvent accessible surface area. *The Journal of Physical Chemistry B*, *105*(21), 5055–5067.

Wang, J., Zheng, X., Yang, Y., Drueckhammer, D., Yang, W., Verkhivker, G., & Wang, E. (2007). Quantifying intrinsic specificity: A potential complement to affinity in drug screening. *Physical Review Letters*, *99*(19), 198101. PMID:18233118

Wang, J.-C., & Lin, J.-H. (2013). Scoring functions for prediction of protein-ligand interactions. *Current Pharmaceutical Design*, *19*(12), 2174–2182. PMID:23016847

Wang, R., Fang, X., Lu, Y., & Wang, S. (2004). The pdbbind database: Collection of binding affinities for protein-ligand complexes with known three-dimensional structures. *Journal of Medicinal Chemistry*, *47*(12), 2977–2980. PMID:15163179

Wang, R., Lai, L., & Wang, S. (2002). Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *Journal of Computer-Aided Molecular Design*, *16*(1), 11–26. PMID:12197663

Wang, R., & Wang, S. (2001). How does consensus scoring work for virtual library screening? an idealized computer experiment. *Journal of Chemical Information and Computer Sciences*, *41*(5), 1422–1426. PMID:11604043

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Xiang, Z., Soto, C. S., & Honig, B. (2002). Evaluating conformational free energies: The colony energy and its application to the problem of loop prediction. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(11), 7432–7437. PMID:12032300

Xue, M., Zheng, M., Xiong, B., Li, Y., Jiang, H., & Shen, J. (2010). Knowledge-based scoring functions in drug design. 1. developing a target-specific method for kinase-ligand interactions. *Journal of Chemical Information and Modeling*, *50*(8), 1378–1386. PMID:20681607

Yan, Z., & Wang, J. (2012). Specificity quantification of biomolecular recognition and its implication for drug discovery. *Scientific Reports*, *2*, 309. PMID:22413060

Yan, Z., & Wang, J. (2015). Optimizing the affinity and specificity of ligand binding with the inclusion of solvation effect. *Proteins: Structure, Function, and Bioinformatics*, 83(9), 1632-1642.

Yan, Z., Zheng, X., Wang, E., & Wang, J. (2013). Thermodynamic and kinetic specificities of ligand binding. *Chemical Science*, *4*(6), 2387–2395.

Yin, S., Biedermannova, L., Vondrasek, J., & Dokholyan, N. V. (2008). Medusascore: An accurate force field-based scoring function for virtual drug screening. *Journal of Chemical Information and Modeling*, *48*(8), 1656–1662. PMID:18672869

Yuriev, E., & Ramsland, P. A. (2013). Latest developments in molecular docking: 2010–2011 in review. *Journal of Molecular Recognition*, *26*(5), 215–239. PMID:23526775

Yusuf, D., Davis, A. M., Kleywegt, G. J., & Schmitt, S. (2008). An alternative method for the evaluation of docking performance: RSR vs RMSD. *Journal of Chemical Information and Modeling*, *48*(7), 1411–1422. PMID:18598022

Zilian, D., & Sotriffer, C. A. (2013). Sfcscore rf: A random forest-based scoring function for improved affinity prediction of protein–ligand complexes. *Journal of Chemical Information and Modeling*, *53*(8), 1923–1933. PMID:23705795

Zou, X., Sun, Y., & Kuntz, I. D. (1999). Inclusion of solvation in ligand binding free energy calculations using the generalized-born model. *Journal of the American Chemical Society*, *121*(35), 8033–8043.

KEY TERMS AND DEFINITIONS

Benchmark: A set of diverse protein-ligand complexes used for testing the scoring functions or docking methods.

Conformational Space: The degrees of freedom involved in the process of placing a ligand onto the protein.

Docking: The process of searching the possible locations for a ligand onto the binding site of a protein.

Sampling: The exploration of the conformational space given by the movements of the docking partners.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Scoring: The measurement of the fitness for the interactions of a ligand with a protein.

Solvation: The process that the water molecules surround the protein and the ligand.

Specificity: The feature of a ligand to recognize its own binding partner and to discriminate against others.

Virtual Screening: Computationally search the library of small molecules to identify the lead compounds which most favorably bind with a target protein.

This work was previously published in Methods and Algorithms for Molecular Docking-Based Drug Design and Discovery edited by Siavoush Dastmalchi, Maryam Hamzeh-Mivehroud, and Babak Sokouti, pages 220-245, copyright year 2016 by Medical Information Science Reference (an imprint of IGI Global).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 37
Online Molecular Docking Resources

Adriana Isvoran
West University of Timisoara, Romania

ABSTRACT

This chapter aims to present the available online resources that are used for protein modeling with accent to online molecular docking resources. SwissDock, MTiAutoDock, and PatchDock online docking tools are described and a few illustrative examples concerning the molecular docking studies for the cytochrom P450 interactions with the fungicide difenoconazole. The results obtained using different servers based on distinct approaches are compared and the advantages and/or disadvantages of every server are illustrated.

INTRODUCTION

In the actual period we assist to the expansion of structural data concerning proteins and protein–ligand complexes with direct implications in the development of the computational strategies to enhance drug discovery. One of these strategies is the virtual screening (VS) technique that allows hits identification with lower costs compared with the traditional experimental high-throughput screening (Gohlke & Klebe, 2002). Virtual screening techniques can be classified into ligand-based and structure-based methods respectively.

The ligand-based method is used when a set of ligand molecules is known but the target's structure is unknown and it may be implemented using several approaches. When the information contained in the set of known ligands is used to build a model of the receptor molecule, the ligand-based method is implemented as pharmacophore modeling (Sun, 2008). The ligand-based approach may be implemented using the quantitative structure activity relationship (QSAR) method that takes into account the connection between chemical structures of some chemical compounds and their biological activities and predicts the activities of new chemicals (Nantasenamat et al, 2009). Another approach is the chemical similarity searching, a method based on the assumption that all compounds in a database that are similar to a query compound have similar biological activity (Willet et al., 1998). When applying this method, a

DOI: 10.4018/978-1-5225-0549-5.ch037

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

database of molecules is screened against one or more known active ligands structures in order to predict their biological activity. The ligand-based virtual screening may also be based on the shape similarity searching, molecules with shape similar to that of the active compound are considered to fit the target's binding site (Rush et al., 2005).

The structure-based drug design involves molecular docking as the most common method. It is based on docking of a candidate compound into the protein target and applying a scoring function to estimate the probability that the ligand will bind to the protein with high affinity (Kroemer, 2007). Computational programs based on diverse algorithms were developed to implement molecular docking studies, molecular docking being an increasingly important tool in actual pharmaceutical researches. From this point of view, it becomes important to evaluate the existing docking programs in order to choose the most suitable docking algorithm for a particular study.

Molecular docking studies may be implemented by using commercial software, free software designed for academic users and also using on-line specialized servers. There are many servers allowing online protein-protein and protein-ligand molecular docking. For example, Click2Drug server (http://www.click2drug.org/) contains a comprehensive list of docking web services, databases and computer-aided drug design tools.

Within this chapter, the use of some online resources designed to protein modeling and online servers allowing molecular docking studies based on different approaches is presented. The description is illustrated with examples of molecular docking computations concerning the prediction of the interactions of the cytochrome P450 with the fungicide difenoconazole (DFC). The results obtained using different servers are compared and discussed in terms of agreement or disagreement concerning the predictions of the receptor-ligand complexes obtained using the docking tools.

BACKGROUND

Online Databases

Last decades are characterized by the development of the internet and increase of the on-line resources. Another characteristic of the last decades is the massive accumulation of data concerning the molecules with therapeutic potential, pesticides and drugs, sequences and structures of the biological macromolecules. These data are now structured and deposited in databases, most of them being freely accessible for educational purposes and they can be used to build models for designing new drug molecules. Existing databases may be classified in databases containing information concerning the small chemical compounds structures and properties, databases comprising information about the sequences and structures of the biological macromolecules and databases relating the information about the chemical compounds with those concerning their molecular targets and pathways. Another more detailed classification of databases that are used for bioinformatics applications considers the following categories: nucleotide sequence databases, genome databases, protein sequence databases, protein structure databases, protein model databases, proteomics databases, nucleic acids databases, carbohydrate structure databases, protein-protein and other molecular interactions, signal transduction pathway databases, metabolic pathway and protein function databases, metabolomic databases and many other specialized databases.

Databases are indispensable resources for online-docking and/or other computational tools used in protein-protein and protein-ligand interactions analysis and drug discovery. A comprehensive view of

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

the chemical databases is provided on the Computational Resources for Drug Discovery server (http://crdd.osdd.net/chemdatabase.php). A few examples of usual chemical libraries databases indispensable for molecular modeling studies are the following:

- ZINC database (http://zinc.docking.org/) provided by the Irwin and Shoichet Laboratories in the Department of Pharmaceutical Chemistry at the University of California and designed for structure-based virtual screening studies (Irwin and Stoichet, 2005; Irwin et al., 2012). It contains over 20 million ligands that may be downloaded in many common file formats (SMILES, mol2, ready-to-dock 3D formats compatible with most docking programs, etc.). The database also permits to extract information concerning the structure, biological activity, physical and chemical properties, name, catalog and Chemical Abstracts Service (CAS) numbers and vendor for every chemical compound. The chemical compound is prepared in multiple protonation states and multiple tautomeric forms and multiple conformations are available for the molecules. There also is the possibility for users to process their own molecules by uploading them to the server.
- PubChem compound database (http://www.ncbi.nlm.nih.gov/pccompound) provides substance information, compound structures and biological activities and is organized as three linked databases: PubChem Substance, PubChem Compound (Bolton et al., 2008) and PubChem BioAssay (Wang et al., 2014). PubChem Substance contains more than 180 million records, PubChem Compound covers more than 63 million structures and PubChem BioAssay comprises more than 1 million BioAssays. PubChem also provides a fast chemical structure similarity search tool.
- ChEMBL database (https://www.ebi.ac.uk/chembldb/) provides information about 1 million small drug-like molecules and 8200 drug targets.
- ChemDB database (http://cdb.ics.uci.edu/) includes information concerning the predicted or experimentally determined physicochemical properties and 3D structures for about 5 million chemicals (Chen et al., 2007).
- ChemSpider database (http://www.chemspider.com/) is owned by the Royal Society of Chemistry and provides access to structures, properties, and associated data integrated with publication information for over 35 million chemical molecules.
- Pesticide databases, such as Pesticide Action Network (PAN) database (http://www.pesticideinfo.org/) containing toxicity and regulatory information for pesticide and Pesticides Properties Database (PPDB, http://sitem.herts.ac.uk/aeru/ppdb/en/) comprising physicochemical, toxicological, ecotoxicological and other related data for about 1150 pesticides, 700 of their metabolites and 100 other related substances and
- Drugs databases: DrugBank database (http://www.drugbank.ca/), Therapeutic Target database (http://xin.cz3.nus.edu.sg/group/ttd/ttd.asp), PharmGKB database (https://www.pharmgkb.org/), Search Tool for Interactions of Chemicals (STITCH, http://stitch.embl.de/), Super Target Database (http://bioinf-apache.charite.de/supertarget_v2/).
- Universal Protein (UniProt) resource (http://www.uniprot.org/) is freely accessible and contains information concerning protein sequences and functions (Jain et al., 2009).The UniProt resource covers three linked databases: UniProt Knowledgebase (UniProtKB), the UniProt Reference Clusters (UniRef), and the UniProt Archive (UniParc).
- Protein Data Bank (PDB, http://www.rcsb.org/pdb/home/home.do) is the single worldwide online resource containing the three dimensional structures of the proteins and nucleic acids and of their complexes with other molecules (Bernstein et al., 1977). These structures are usually ob-

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

tained through X-ray crystallography and/or NMR studies. PDB contains structures for biological molecules belonging to diverse organisms and it is updated weekly. On 11 of October 2015 this database contains 112 722 structures. The server housing the PDB resource also provides tools and additional information for researchers and students in structural biology, molecular biology, computational biology and other related fields.

- PDBsum (https://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPage.pl?pdbcode=index.html) is a pictorial database that offers a quick overview of the contents of each structural file deposited in the Protein Data Bank illustrating the molecule(s) contained in the structural file with a schematic diagram of their interactions (de Beer et al., 2014). In the case of proteins, for every PDB entry, the information deposited in the PDBSum is ranged on several pages and it contains: summary information (image of the structure and the molecules therein, enzyme reaction diagram), wiring diagram, topology diagram, residue conservation, description of ligands contained in the structural file, schematic representations of protein-protein interfaces and the residue-residue interactions made across them, top ten clefts in the surface of the protein classified by volume with the bound ligands shown and links to external databases for further information.

There also are other specialized online databases that are used for particular bioinformatics applications but it is not the aim of this chapter to present all available databases in the field.

Protein Modeling Resources

Due to advances in computer technology, the massive amount of available biological data and progresses in modeling methods development, protein modeling is playing a more and more important role in many fields of biological and biochemical sciences. Protein modeling tools refer to the prediction and/or analysis of the protein structure in correlation with its function and its underlying mechanism and may also provide insight to design experiments and suggest possible clues for drug design and discovery. There are online resources available for protein modeling and they usually refer to (Xu et al., 2000):

1. Protein sequence comparison including sequences alignment, sequence-based protein family classification, phylogenetic classification;
2. Protein sequence annotation including prediction of hydrophobic and/or flexibility profiles, of the secondary structures of proteins, of the transmembrane regions, active or signaling sites;
3. Three-dimensional structure prediction by homology modeling;
4. Protein structure analysis including structure visualization and geometry analysis, structure comparison and classification, surface analysis with pockets and cavities detection, electrostatics and hydrophobic surface coloring;
5. Molecular dynamics simulation and molecular docking aiming the protein-protein, protein-ligands and other protein interactions analysis and characterization.

Protein structure analysis, structures comparison and classification require molecular graphics visualization tools that enable the visualization of the structures that are encoded by the atomic coordinate files expressed in different formats and manipulation of these images to observe the molecules from various perspectives. There are several open source visualization tools that are freely download and used through a wide variety of operating systems (Table 1).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 1. Free molecular visualization tools

Visualization Tool	URL	Reference
UCSF Chimera	http://www.cgl.ucsf.edu/chimera/	Pettersen et al., 2004
PyMol	https://www.pymol.org/	http://sourceforge.net/projects/pymol/
Swiss PDB Viewer	http://spdbv.vital-it.ch/	Guex and Peitsch, 1997
Jmol	http://jmol.sourceforge.net/	http://www.jmol.org/
Rasmol	https://www.umass.edu/microbio/rasmol/	https://www.umass.edu/microbio/rasmol/distrib/rasman.htm
Visual molecular dynamics (VMD)	http://www.ks.uiuc.edu/Research/vmd/	Humphrey et al., 1996

Proteins perform their biological functions by interacting with other molecules (substrates, antigens, nucleic acids, peptides, proteins, etc). These interactions occur in specific regions on the protein surface that are either concavities on the protein surface (pockets) either regions enveloped by the protein surface (cavities). Therefore, it is of interest to identify pockets and cavities in protein structures. Especially, the detection of pockets and cavities is a prerequisite for identification and characterization of pockets and/or cavities presents on the protein surface are important tasks. This information can provide insight into biological functions and reaction mechanisms of proteins and it is also a requirement for protein–ligand docking and structure-based drug design. There are a few online resources destined to the detection and characterization of pockets and/or cavities on the protein surface, such as:

- Fpocket (http://mobyle.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py?form=fpocket2#forms::fpocket), a computational tool that enables the identification and characterization of the potent binding sites on the protein surface (Schmidtke et al. 2010).
- Computed Atlas of Surface Topography of proteins (CASTp, http://sts.bioe.uic.edu/castp/index.php) that allows the identification and measurements of surface accessible pockets and of interior inaccessible cavities for proteins and nucleic acids (Dundas et al., 2006).
- DEPTH (http://mspc.bii.a-star.edu.sg/tankp/intro.html) is a web server used for structural analyses of protein structures that permits to detect potential small-molecule ligand-binding cavities and compute their sizes and to predict the pKa of ionizable residues in proteins (Tan et al., 2013).
- Roll (http://altair.sci.hokudai.ac.jp/g6/Research/POCASA_e.html.) is an algorithm that can predict binding sites by detecting pockets and cavities of proteins with a rolling sphere (Yu et al., 2010). Roll is implemented in the program POCASA that can automatically predict the ligand binding sites of proteins. This program uses different probe spheres to determine the pockets and it makes the program adaptable for various ligands and proteins.
- *3D-SURFER* (http://dragon.bio.purdue.edu/3d-surfer) is a web-based tool destined to an interactive protein surface analysis, comparison and characterization of proteins at the global and local levels based on their surface shape (La et al., 2009). Concerning the protein surface analysis, the server also identifies and visualizes local features, such as pockets, protrusions, hollow spaces and flat regions. It also integrates identification of protein surface properties (electrostatic potentials, hydrophobicity and conservation) allowing detailed analysis designed to assist investigating function of proteins.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

There also are resources concerning the protein spatial structure prediction that is one of the most important goals pursued by computational biochemistry especially for the drug design and biotechnological studies. The protein 3D structure prediction tools employ either homology based methods or *ab initio* methods. Homology based protein structure prediction requires the identification of homologs that can be used as templates for modeling with a significant amino acid sequence similarity of the model to the target sequence, this search involving the use of specialized computational tools (Wallner & Elofsson, 2005). Some online resources designed to the structure homology modeling of proteins are presented in Table 2.

Ab initio structure prediction methods are used when there are no available structural models for the investigated protein, the structure is predicted from sequence alone and they are usually designed to small proteins (Bonneau and Baker, 2001).

Molecular Docking Resources

Besides the great diversity of the bioinformatics tools, the molecular docking is one of the computational tools that help us to understand the interactions between biologically relevant molecules (proteins, nucleic acids, carbohydrates, lipids) at the atomic level, as well as to clarify fundamental biochemical processes, especially those based on the protein-protein and protein–ligands interactions.

Molecular docking is a computational technique focusing on prediction of the binding mode(s) (BMs) of a ligand with a protein (Morris & Lim-Wilby, 2008). It aims of finding an optimized conformation for both the protein and ligand and their relative orientations such that the free energy of the protein-ligand complex is minimum. There are two approaches used in molecular docking studies: one is the geometry-based approach that is centered on the surface complementarity of the ligand and protein (Shoichet et al., 2004) and the other approach is energetic-based and consists on simulating the docking process and computing ligand-protein pairwise interactions energies (Feig et al., 2004). Every approach has both advantages and limitations (Meng et al., 2011).

Performing molecular docking using the surface complementarity approach means to obtain the molecular surfaces of both the protein and ligand (knowing their spatial structures) and to find the rigid transformations such as the tow objects have a good match, i.e. they share the maximum contact area between their molecular surfaces without overlapping regions (Norel et al., 1994). Usually, the entire molecular surfaces are utilized in the docking process, not only the location of the active center of the

Table 2. Free protein structure homology modeling tools

Structure Homology Tool	URL	Reference
MODELLER	http://salilab.org/modeller/	**Eswar et al, 2006**
SWISS-MODEL	http://swissmodel.expasy.org/interactive	Biasini et al, 2014
HHpred	http://toolkit.tuebingen.mpg.de/hhpred	Söding, 2005
Geno3D	http://geno3d-pbil.ibcp.fr/cgi-bin/geno3d_automat.pl?page=/GENO3D/geno3d_home.html	Combet et al., 2002
ESyPred3D	http://www.unamur.be/sciences/biologie/urbm/bioinfo/esypred	Lambert et al., 2002
PHYRE2	http://www.sbg.bio.ic.ac.uk/ phyre2	Kelley et al., 2015

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

target protein, or some particular residues or atoms participating in the binding in either the protein or the ligand. There is a matching of every piece of surface of the ligand with every piece of surface of the target (Norel et al., 1999). Most docking algorithms use two kinds of molecular descriptors for the molecules involved, structural complementarity and binding complementarity descriptors (Oloff et al., 2006). Structural complementarity descriptors include solvent-accessible surface area, overall shape and geometric constraints between atoms in the protein and ligand and binding complementarity descriptors consider hydrogen bonds, hydrophobic contacts and van der Waals interactions. In order to limit the number of obtained binding modes, additional constraints could be used as additional filters, for example electrostatic interactions. The surface complementarity between the ligand and its target is implemented in two ways: grid-based (Gabb et al., 1997) and pairwise (Chen & Weng, 2003) shape complementarity respectively. Grid-based shape complementarity is computed by considering the grid points on at the binding interface (Gabb et al., 1997) and pairwise shape complementarity is computed by considering the total number of atom pairs between the protein and the ligand within a distance cutoff (Chen and Weng, 2003).

In the case of simulation of the docking process, the ligand and the protein are separated by a physical distance. This approach usually considers proteins as rigid bodies and the ligand is allowed to explore their conformational space. The ligand must find its position into the protein's active center using a certain number of changes in its conformational space (translations, rotations, torsions).The generated conformations of the ligand are docked consecutively into the protein, and a molecular dynamics (MD) simulation is performed. During the MD process, each of the ligand changes induces an energetic cost of the system and the system's total energy is calculated taking into account every change and the energies values determined are used for ranking the overall scoring (Halperin et al., 2002; Feig et al., 2004). Simulation of the docking process has the advantage that allows to easily consider the ligand flexibility, but it is computationally more expensive involving potentially hundreds of MD runs.

The implementation of docking algorithms implies that the three dimensional structures of both the ligand and its target are known and there are multiple formats for the files describing the three dimensional structures (Dalby et al., 1992). In the case of the target protein, the three dimensional structural files are obtained from the PDB. Structural data have a *PDB format*, a standard representation of the macromolecular structural data containing atomic coordinates. The PDB format (with the typical file extension.pdb) is a fixed-column-width format commonly used to express the proteins structures, but it may also be used for other types of molecules. This format is read and also written by many programs used to visualize and analyze the molecular structures. Some docking algorithms use as input for the target molecule its structural file in a *pdb* format, but others use the *mol2* format for this purpose. The *mol2* (with the typical file extension.*mol2*) format contains information about the atoms, bonds, connectivity and atoms coordinates, also being a standardized format that some modeling programs can read and write. Another type of file format is MDL Molfile with the typical extension.*mol*. It contains information concerning atoms coordinates, bonds and connectivity for a molecule and numerous computational applications are able to read it (Dalby et al., 1992).

In the case of small molecules (ligands), there are several formats offered by the chemical libraries. The common format used to encode small chemical compounds is the Simplified Molecular Input Line Entry Specification (SMILES) format. This format includes information concerning the connectivity but do not include 2D or 3D coordinates of the atoms and the hydrogen atoms are not represented (Weininger, 1988). Every atom is represented by its element symbol, a double bond is represented by the symbol "=", a triple bond by the symbol"#", branching is represented by "()" and rigs are indicated

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

by pairs of digits. Another format widely used for the small chemical compounds is the Structure Data Format (SDF), a text file commonly used to represent multiple chemical structure records and associated data fields (Dalby et al., 1992).

Taking into account the diverse formats used to represent a molecule and that every computational tool used for viewing and editing molecular structures or for molecular modeling is capable to read and write only one (or a reduced number) of the file format types, format converters have been developed. OpenBabel (http://openbabel.org/wiki/Main_Page) is one of the commonly used online freely available tools specifically designed for converting between file formats (O'Boyle et al., 2011). It can read 88 formats and can write 89 formats. Another online format converter is PRODRG (http://davapc1.bioch.dundee.ac.uk/cgi-bin/prodrg). It reads many types of coordinates formats (pdb, mdl, mol2) or two-dimensional formats and generates a variety of topologies for use with molecular modeling tools, as well as energy-minimized coordinates in a variety of formats (Schüttelkopf & van Aalten, 2004).

There are multiple servers allowing online protein-ligand and protein-protein molecular docking. Table 3 reveals a short presentation of the most known online docking resources.

The use of PatchDock, SwissDock and MTiAutoDock servers is further illustrated.

Table 3. Online molecular docking servers

Molecular Docking Tool	URL	Short Description	Reference
DOCK Blaster	http://blaster.docking.org	DOCK Blaster is an online server used for the prediction of the ligand-protein complexes.	Irwin et al., 2009
Molecular Docking Server	http://www.dockingserver.com/web	DockingServer enables docking calculations and results evaluation. It can be used for docking of single ligands as well as for docking of ligand libraries to target proteins.	Bikadi and Hazai, 2009
ZDOCK	http://zdock.umassmed.edu	ZDOCK is a docking algorithm that searches all possible binding modes between the two proteins and evaluates each pose using an energy-based scoring function.	Pierce et al., 2014
Rosseta	http://rosie.rosettacommons.org/	Rosetta is an online tool used for docking small molecules into proteins.	Combs et al., 2013
PatchDock	http://bioinfo3d.cs.tau.ac.il/PatchDock/)	PatchDock is a webserver performing structure prediction for protein-small ligand complexes using a geometry based molecular docking algorithm.	Schneidman-Duhovny et al., 2005
SwissDock	http://www.swissdock.ch/	SwissDock accomplishes molecular docking predictions based on computing ligand-protein pairwise interactions energies.	(Grosdidier et al., 2011)
MTiAutoDock	http://mobyle.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py#forms::MTiAutoDock	MtiAutoDock permits docking of small ligands into proteins cavities based on computing ligand-protein pairwise interactions energies.	Labbe et al., 2015
HexServer	http://hexserver.loria.fr/	HexServer is an interactive protein docking and molecular superposition program. It performs FFT-based rigid-body docking predictions.	Macindoe et al, 2010
HADDOCK (High Ambiguity Driven protein-protein DOCKing)	http://haddock.science.uu.nl/services/HADDOCK/	HADDOCK deals with modelling problems including protein-protein, protein-nucleic acids and protein-ligand complexes.	de Vries et al., 2010

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

SOLUTIONS AND RECOMMENDATIONS

When performing online molecular docking studies information concerning the ligand and protein structures is needed. The ligand structure and properties may be retrieved from ZINC (Irwin and Stoichet, 2005), PubChem (Bolton et al., 2008), ChEmbl, ChemDB (Chen et al., 2007), ChemSpider or other database. Depending on the format delivered by the chemical database and the format read by the docking algorithm, a format converter can be necessary. The structure of the target protein is retrieved from the Protein Data Bank (Bernstein et al., 1977). For molecular docking studies, the structures of protein-ligand complexes are considered if available and the ligand is removed when the receptor molecule is prepared for docking.

PatchDock server uses an algorithm that is based on finding the transformations that produce a local good geometric shape complementarity including the local individualities of the docked molecules with complementary features. The transformations are classified using a scoring function tacking into account both the geometric fit and atomic desolvation energy, the redundant solutions being rejected by applying a root mean square deviation clustering (RMSD). The output of PatchDock tool is a list of candidate complexes between the protein and the ligand molecule that were specified by user, the complexes being sorted according to the geometric shape complementarity score. For every obtained complex, the output contains the approximate interface area, the atomic contact energy (ACE) and the 3D transformations (3 rotational angles and 3 translational parameters) applied on the ligand molecule. To refine the PatchDock predictions, theFireDock webserver (http://bioinfo3d.cs.tau.ac.il/FireDock/) is used. It also provides the global energy of each protein-ligand complex predicted by PatchDock software (Andrusier et al., 2007). The user must upload the structural files of the protein (receptor) and ligand in the PDB format or provide their PDB codes entry. As crystal structures of proteins sometimes contains ions, water molecules and/or small ligands, preparation of the structures for docking studies keeping the functional molecules is usually recommended. The user must also provide an e-mail address to obtain the link to the docking results and the value of the RMSD used for the clustering, the recommended value being 1.5 Å for protein-small molecule docking. PatchDockserver uses different sets of parameters for distinct types of complexes (enzyme-inhibitor, antibody-antigen and protein-small ligand) and the user must specify the type of the complex. If the user does not know the type of the complex, the default complex type may be used. There also are some advanced options that are used when some information concerning the potential binding sites in the protein is known, a list of residues in an uploaded file may be specified.

The use of PatchDock server is illustrated here for the prediction of the interaction between the human cytochrome P450 subfamily 2C8 (CYP2C8) with the pesticide difenoconazole (DFC). The cytochrome P450 enzymes (CYPs) are proteins with critical roles in xenobiotic metabolism, and the CYP2C subfamily is involved in the metabolism of about 49% of widely used pesticides (Abass et al., 2011). Low levels of pesticides from residues in food and drinking water, respectively the occupational exposure may cause health problems in humans and the assessment of the pesticides interactions with human CYP2C8 is important as it is one of the main hepatic drug-metabolizing enzymes. There are 5 structural files in the PDB for the human CYP2C8 protein, their codes entry being: 1PQ2 for the complex of CYP2C8 with palmitic acid (Schoch et al, 2004), 2NNH, 2NNI, 2NNJ and 2VN0 for the complexes with 9-cis-retinoic acid, montelukast, felodipine, and troglitazone respectively (Schoch et al., 2008). In our study we have used the crystal structure of the CYP2C8 in complex with the inhibitor montelukast, code entry 2NNI. The inhibitor molecule, the palmitic acid molecule and the SO_4 ion have been removed, the heme group

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

was kept and the resulting file has been prepared for docking using the UCSF Chimera computational tool. The output of the PatchDock server for the 20 best solutions is shown in the Table 4.

The solutions presented in Table 4 are ranked upon the shape complementarity score. The best solution predicted by PatchDock server (result.1.pdb) is illustrated in the Figure 1. Figure 1 reveals that DFC (presented as full surface) does not bind in the same place as montelukast (mesh surface), the known inhibitor of the protein.

The best solutions obtained using PatchDock have been refined using FireDock on-line computational tool and the best 10 refined solutions are presented in Table 4. This table contains the new ranking after refining, the global energy, the attractive and repulsive van der Waals (VdW) energies, the atomic contact energy (ACE) and hydrogen bonds (HB). This table reflect that, after refining, the best solution is considered solution number 8 predicted by PatchDock, this solution being illustrated in the Figure 2. Table 5 reflects that solution number 8 has the highest values of the atomic contact energy.

Figure 2 illustrates the importance of refining results when using molecular docking based on the shape complementarity of protein and ligand. After refining, the best solution reflects that DFC is situated almost in the same place as the known inhibitor and it suggests that DFC may also have an inhibitory effect on the human CYP2C8 protein.

SwissDock is a webserver (http://www.swissdock.ch/) performing molecular docking predictions based on computing ligand-protein pairwise interactions energies (Grosdidier et al., 2011). This server

Table 4. The 20 best solutions predicted by PatchDock for the DFC binding to the human CYP2C8

Solution No	Score	Interface Area (Å^2)	ACE (kcal/mol)	Transformation	PDB File of the Complex
1	5694	664.50	-163.40	2.22 0.17 -2.97 52.79 13.42 -11.67	result.1.pdb
2	5540	678.40	-216.28	2.80 -0.22 -2.05 48.02 13.65 -19.35	result.2.pdb
3	5448	652.50	-174.57	-0.86 -0.04 0.11 46.88 12.71 -21.72	result.3.pdb
4	5430	626.50	-160.43	-1.17 0.03 1.43 49.62 10.23 -25.82	result.4.pdb
5	5372	619.20	-164.94	-0.92 0.16 0.37 48.34 11.88 -25.01	result.5.pdb
6	5244	691.50	-338.52	0.65 -0.13 -2.15 52.45 -0.47 -20.01	result.6.pdb
7	5242	663.10	-234.23	2.59 -0.15 -2.41 49.72 13.33 -19.86	result.7.pdb
8	5228	651.10	-257.81	1.37 0.01 0.55 47.03 11.80 -24.13	result.8.pdb
9	5222	599.10	-139.50	2.39 -0.21 -3.04 50.28 14.14 -12.68	result.9.pdb
10	5220	608.90	-171.76	-1.05 -0.04 -0.30 47.38 12.81 -23.67	result.10.pdb
11	5184	662.70	-245.84	2.27 0.17 -1.86 51.05 13.70 -20.15	result.11.pdb
12	5166	630.50	-199.85	1.84 -0.56 1.08 48.66 12.95 -18.09	result.12.pdb
13	5166	613.50	-228.08	-1.46 0.46 -2.90 48.01 16.30 -27.79	result.13.pdb
14	5158	627.50	-196.07	-1.16 0.28 0.09 48.16 11.22 -23.28	result.14.pdb
15	5156	648.10	-166.58	-1.36 -0.75 -1.35 47.28 16.26 -16.15	result.15.pdb
16	5144	621.70	-170.42	2.32 -0.02 2.92 51.05 12.82 -11.72	result.16.pdb
17	5134	636.10	-172.31	-1.15 -0.14 -1.10 46.94 15.56 -20.76	result.17.pdb
18	5128	617.40	-170.80	-0.96 0.61 0.43 50.64 12.64 -25.90	result.18.pdb
19	5124	649.60	-211.24	-1.40 0.77 0.18 50.34 10.92 -24.16	result.19.pdb
20	5122	643.40	-191.75	2.84 -0.54 -2.29 47.81 13.87 -16.78	result.20.pdb

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 1. The binding poses for the difenoconazole to the human CYP2C8 predicted using PatchDock server. The protein is shown as backbone, the difenoconazole in revealed as full surface, the inhibitor montelukast is shown as mesh surface.

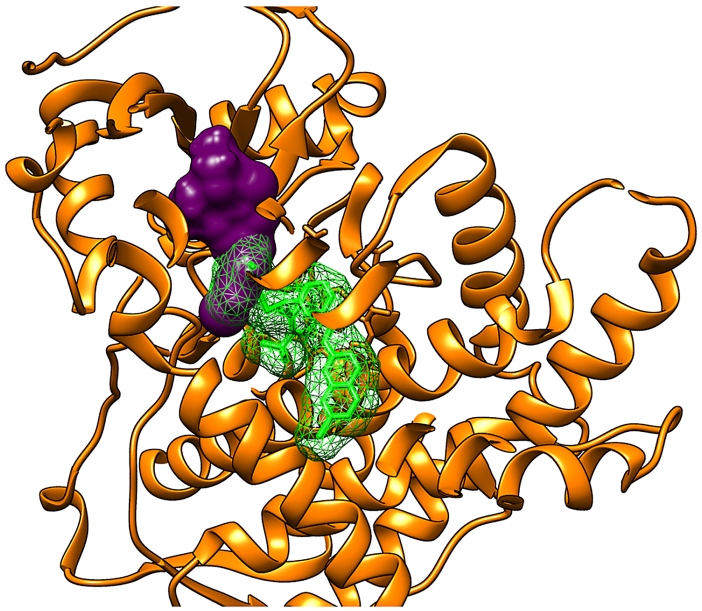
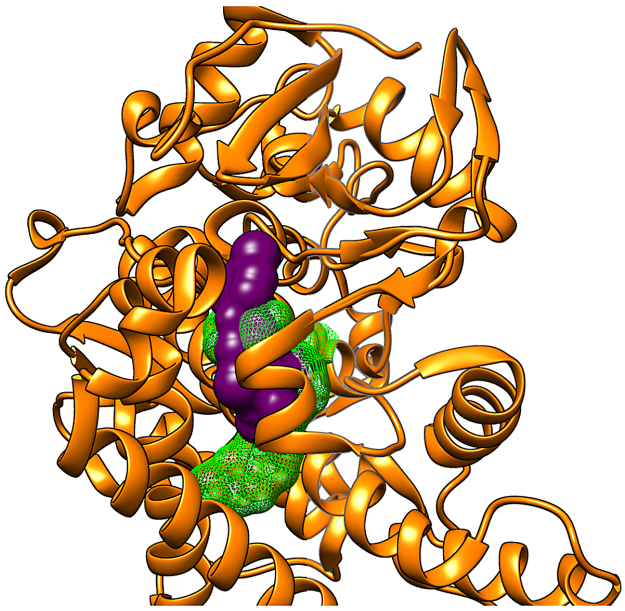


Figure 2. The binding poses for the difenoconazole to the human CYP2C8 predicted using PatchDock server and refined using FireDock server. The protein is shown in backbone, the difenoconazole in revealed as full surface, the inhibitor montelukast is shown in mesh surface.



EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 5. The 10 best refined solutions predicted by FireDock for the DFC binding to the human CYP2C8

Rank	Solution Number	Global Energy (kcal/mol)	Attractive VdW (kcal/mol)	Repulsive VdW (kcal/mol)	ACE (kcal/mol)	HB
1	8	-43.02	-17.47	7.62	-14.42	0.00
2	7	-35.57	-13.64	1.35	-10.23	0.00
3	3	-35.54	-19.93	11.40	-9.74	0.00
4	1	-35.52	-16.62	5.01	-9.43	0.00
5	2	-32.98	-14.52	4.22	-9.19	0.00
6	9	-31.36	-15.23	6.90	-8.97	0.00
7	5	-30.95	-13.65	4.05	-8.91	0.00
8	4	-30.24	-14.12	5.73	-8.96	0.00
9	10	-28.44	-14.98	9.11	-8.35	0.00
10	6	-2.49	-9.90	5.36	3.92	0.00

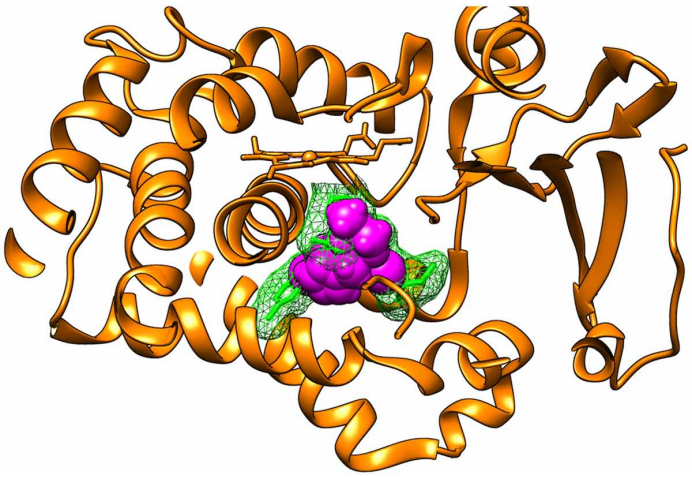
uses the EADock (Grosdidier et al., 2007) and EADockDSS (Grosdidier et al, 20111 b) algorithms to compute the interacting energies and outcomes the most favorable position and orientation of a ligand on a protein surface, i.e. the binding modes (BMs). The interaction energy is expressed as the FullFitness score computed by averaging the 30% most favorable effective energies of its elements. The effective energy considers the total energy of the system and the solvation energy. The obtained BMs are clustered and clusters are ranked by averaging the FullFitness scores of their elements. The protein structure can be specified by its code from the Protein Data Bank or by uploading the structural file. As mentioned above, it is recommended to upload a structural file of the protein prepared for docking studies. The ligand can be specified either by its identifier from the ZINC database or by uploading its structure files. There are three possibilities to select the docking type: very fast, fast and accurate. Three docking parameter presets can be selected: very fast, fast and accurate. They differ by the number of sampled BMs, the number of minimization steps and the number of predicted BMs. The user has the possibility to define or not the region of interest for docking. If there are not information concerning the binding site or the site of interest, the whole protein structure is considered during the docking process and the region of interest is not defined. If the binding pocket is known, the docking can be restricted to this region of the protein structure by defining the region of interest. The user may specify an email address in the submission form and a link to the docking result web page is sent to this address when the docking process in finished. The predicted BMs may be visualized directly within the web browser or they may be downloaded for analyzed using Chimera software.

The SwissDock server is also used to predict the interaction between the CYP2C8 protein and the pesticide difenoconazole. In this purpose, the same structural files of the protein and ligand are used as previously explained for the PatchDock webserver and the region of interest is not defined. SwissDock webserver predicts 38 BMs, the best pose of the DFC binding to the CYP2C8 being presented in Figure 3. The fulfitness score for this pose is -2287.30 kcal/mol, the free energy of binding being -8.28kcal/mol.

The position of the predicted best pose (having the highest predicted fulfitness score and binding free energy) is compared to that of the known inhibitor of the CYP2C8, montelukast, in the crystallographic structure of the CYP2C8-montelukast complex. Figure 2 illustrates that DFC is superposed to the montelukast molecule.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 3. The binding poses for the difenoconazole to the human CYP2C8 predicted by SwissDock server. The protein is shown as backbone, the difenoconazole in revealed as full surface, the inhibitor montelukast is shown as mash surafce and the heme group in sticks.



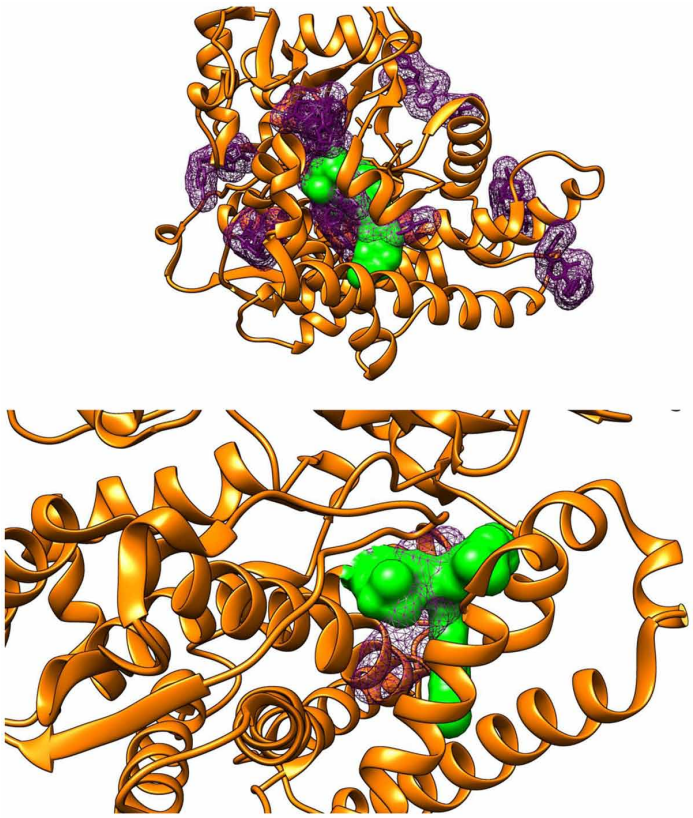
MTiAutoDock (http://mobyle.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py#forms::MTiAutoDock) is a recent designed webserver permitting docking of small ligands into proteins cavities and it is also based on computing ligand-protein pairwise interactions energies (Labbe et al., 2015). It uses the AutoDock4.2 algorithm to generate orientations/conformations of the compound (Morris et al, 2009). The webserver interface allows to the user to upload the ligand molecule in MOL2 format. For the protein, the user may provide the structure in PDB or MOL2 formats. When the protein is provided in PDB format, the structure is cleaned and all heteroatoms are removed and the hydrogen atoms are added. When the heteroatoms (or some of them) must be kept in the structure, the protein structure need to be prepared manually and uploaded in the MOL2 format. MTiAutoDock allows automated blind docking of ligands or ligands docking into a binding site defined by user and outputs the bound conformations and binding energies of ligands with their targets. Up to 10 ligands may be used for docking. The computation of the binding free energy is based on a semiempirical scoring function that allows incorporation of intramolecular energies and of charge based desolvation energy (Morris et al., 2009). When docking predictions are finished, the webserver allows the download and/or visualization in 3D of the 10 top scored poses for the docked ligand(s). MTiAutoDock also allows molecular docking studies with multiple ligands.

For the prediction of the interaction of CYP2C8 and DFC, MTiAutoDock server outcomes, for blind docking, 10 poses, but 6 of them are not predicted in the binding site as revealed by Figure 4A. The pose having the highest free energy of binding, -7.24 kcal/mol, is presented in Figure 4B.

There is a consensus for predicted binding region and free energy of binding values for the protein-difenoconazole complexes obtained using the three docking servers consolidating the prediction of the CYP2C8-fungicide interaction. All molecular docking servers predict that the fungicide may bind to the active site of the protein and that the binding region corresponds to that of the known inhibitor, montelukast, suggesting the inhibitory potential of the fungicide.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 4. The predicted binding poses by MTiAutoDock server for the difenoconazole to the human CYP2C8 (A) and the pose having the highest free energy of binding (B). The protein is shown as backbone, the difenoconazole in revealed as mesh surface, the inhibitor montelukast is shown in as full surface and the heme group in sticks.



FUTURE RESEARCH DIRECTIONS

On-line facilities for performing bioinformatics analysis increase continuously. We assist during the last years to the considerable growth of the number of online resources and also of the accumulation of high-quality data describing structures and affinities of protein-drug ligand complexes in publicly available databases. These data are very important for development, validation, and benchmarking of drug docking algorithms.

CONCLUSION

This chapter presents online protein modeling and molecular docking facilities that can be successfully employed for distinct studies concerning drug design and biotechnology.

The online bioinformatics resources are important for research in all fields of life sciences and they are developing very fast. In the case of teaching, the online bioinformatics resources are important for

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

both teachers and students because they develop interest to the students by facilitating understanding advanced biology studies and they promote quality human resource to undertake challenging research in the field of modern biology and biochemistry.

New online resources for in silico drug screening speed drug discovery and reduce drug development costs by allowing concentrating experimental efforts on the most favorable candidate compounds.

REFERENCES

Abass, K., Turpeinen, M., Rautio, A., Hakkola, J., & Pelkonen, O. (2011). Metabolism of Pesticides by Human Cytochrome P450 Enzymes In Vitro – A Survey. In Insecticides - Advances in Integrated Pest Management (pp 165-194). Rijeka: InTech.

Andrusier, N., Nussinov, R., & Wolfson, H. J. (2007). FireDock: Fast interaction refinement in molecular docking. *Proteins*, *69*(1), 139–159. doi:10.1002/prot.21495 PMID:17598144

Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. E. Jr, Brice, M. D., Rodgers, J. R., & Tasumi, M. et al. (1977). The Protein Data Bank: A Computer-based Archival File For Macromolecular Structures. *Journal of Molecular Biology*, *112*(3), 535–542. doi:10.1016/S0022-2836(77)80200-3 PMID:875032

Biasini, M., Bienert, T., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., & Schwede, T. et al. (2014). SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*, *42*(1), W252–W258. doi:10.1093/nar/gku340 PMID:24782522

Bikadi, Z., & Hazai, E. (2009). Application of the PM6 semi-empirical method to modeling proteins enhances docking accuracy of AutoDock. *Journal of Chemical Information*, *1*, 15. PMID:20150996

Bolton, E., Wang, Y., Thiessen, P. A., & Bryant, S. H. (2008). PubChem: Integrated Platform of Small Molecules and Biological Activities. In R. A. Wheeler & D. C. Spellmeyer (Eds.), *Annual Reports in Computational Chemistry* (pp. 217–241). Oxford, UK: Elsevier. doi:10.1016/S1574-1400(08)00012-1

Bonneau, R., & Baker, D. (2001). Ab initio protein structure prediction: Progress and prospects. *Annual Review of Biophysics and Biomolecular Structure*, *30*(1), 173–189. doi:10.1146/annurev.biophys.30.1.173 PMID:11340057

Chen, J. H., Linstead, E., Swamidass, S. J., Wang, D., & Baldi, P. (2007). ChemDB update-full-text search and virtual chemical space. *Bioinformatics (Oxford, England)*, *23*(17), 2348–2351. doi:10.1093/bioinformatics/btm341 PMID:17599932

Chen, R., & Weng, Z. (2003). A novel shape complementarity scoring function for protein-protein docking. *Proteins*, *51*(3), 397–408. doi:10.1002/prot.10334 PMID:12696051

Combet, C., Jambon, M., Deléage, G., & Geourjon, C. (2002). Geno3D: Automatic comparative molecular modelling of protein. *Bioinformatics (Oxford, England)*, *18*(1), 213–214. doi:10.1093/bioinformatics/18.1.213 PMID:11836238

Combs, S. A., Deluca, S. L., DeLuca, S. H., Lemmon, G. H., Nannemann, D. P., Nguyen, E. D., & Meiler, J. et al. (2013). Small-molecule ligand docking into comparative models with Rosetta. *Nature Protocols*, *8*(7), 1277–1298. doi:10.1038/nprot.2013.074 PMID:23744289

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Dalby, A., Nourse, J. G., Hounshell, W. D., Gushurst, A. K. I., Grier, D. L., Leland, B. A., & Laufer, J. (1992). Description of several chemical structure file formats used by computer programs developed at Molecular Design Limited. *Journal of Chemical Information and Modeling*, *32*(3), 244–255. doi:10.1021/ci00007a012

de Beer, T. A. P., Berka, K., Thornton, J. M., & Laskowski, R. A. (2014). PDBsum additions. *Nucleic Acids Research*, *42*(D1), D292–D296. doi:10.1093/nar/gkt940 PMID:24153109

de Vries, S. J., van Dijk, M., & Bonvin, A. M. J. J. (2010). The HADDOCK web server for data-driven biomolecular docking. *Nature Protocols*, *5*(5), 883–897. doi:10.1038/nprot.2010.32 PMID:20431534

Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y., & Liang, J. (2006). CASTp: Computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Research*, *34*(Web Server), W116–W118. doi:10.1093/nar/gkl282 PMID:16844972

Eswar, N., Marti-Renom, M.A., Webb, B., Madhusudhan, M.S., Eramian, D., Shen, M., Pieper, U., & Sali. A. (2006). Comparative Protein Structure Modeling With MODELLER. *Current Protocols in Bioinformatics*, *15*, 5.6.1-5.6.30.

Feig, M., Onufriev, A., Lee, M. S., Im, W., Case, D. A., & Brooks, C. L. (2004). Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. *Journal of Computational Chemistry*, *25*(2), 265–284. doi:10.1002/jcc.10378 PMID:14648625

Gabb, H. A., Jackson, R. M., & Sternberg, M. J. (1997). Modelling protein docking using shape complementarity, electrostatics and biochemical information. *Journal of Molecular Biology*, *272*(1), 106–120. doi:10.1006/jmbi.1997.1203 PMID:9299341

Gohlke, H., & Klebe, G. (2002). Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angewandte Chemie International Edition*, *41*(15), 2644–2676. doi:10.1002/1521-3773(20020802)41:15<2644::AID-ANIE2644>3.0.CO;2-O PMID:12203463

Grosdidier, A., Zoete, V., & Michielin, O. (2007). EADock: A multiobjective evolutionary framework for the docking of small molecules into protein active site. *Proteins: Structure, Function, and Bioinformatics*, *67*(4), 1010–1025. doi:10.1002/prot.21367

Grosdidier, A., Zoete, V., & Michielin, O. (2011). SwissDock - a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Research*, *39*(suppl), W270–W277. doi:10.1093/nar/gkr366 PMID:21624888

Grosdidier, A., Zoete, V., & Michielin, O. (2011). SwissParam: A fast force field generation tool for small organic molecule. *Journal of Computational Chemistry*, *32*(11), 2359–2368. doi:10.1002/jcc.21816 PMID:21541964

Guex, N., & Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis*, *18*(15), 2714–2723. doi:10.1002/elps.1150181505 PMID:9504803

Halperin, I., Ma, B., Wolfson, H., & Nussinov, R. (2002). Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins*, *47*(4), 409–443. doi:10.1002/prot.10115 PMID:12001221

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD - Visual Molecular Dynamics. *Journal of Molecular Graphics*, *14*(1), 33–38. doi:10.1016/0263-7855(96)00018-5 PMID:8744570

Irwin, J. J., & Shoichet, B. K. (2005). Zinc – a free database of commercially available compounds for virtual screening. *Journal of Chemical Information and Modeling*, *45*(1), 177–182. doi:10.1021/ci049714+ PMID:15667143

Irwin, J. J., Shoichet, B. K., Mysinger, M. M., Huang, N., Colizzi, F., Wassam, P., & Cao, Y. (2009). Automated Docking Screens: A Feasibility Study. *Journal of Medicinal Chemistry*, *52*(18), 5712–5720. doi:10.1021/jm9006966 PMID:19719084

Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S., & Coleman, R. G. (2012). ZINC: A Free Tool to Discover Chemistry for Biology. *Journal of Chemical Information and Modeling*, *52*(7), 1757–1768. doi:10.1021/ci3001277 PMID:22587354

Jain, E., Bairoch, A., Duvaud, S., Phan, I., Redaschi, N., Suzek, B. E., & Gasteiger, E. et al. (2009). Infrastructure for the life sciences: Design and implementation of the UniProt website. *BMC Bioinformatics*, *10*(1), 136. doi:10.1186/1471-2105-10-136 PMID:19426475

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, *10*(6), 845–858. doi:10.1038/nprot.2015.053 PMID:25950237

Kroemer, R. T. (2007). Structure-based drug design: Docking and scoring. *Current Protein & Peptide Science*, *8*(4), 312–328. doi:10.2174/138920307781369382 PMID:17696866

La, D., Esquivel-Rodriguez, J., Venkatraman, V., Li, B., Sael, L., Ueng, S., & Kihara, D. et al. (2009). 3D-SURFER: Software for high throughput protein surface comparison and analysis. *Bioinformatics (Oxford, England)*, *25*(21), 2843–2844. doi:10.1093/bioinformatics/btp542 PMID:19759195

Labbé, C. M., Rey, J., Lagorce, D., Vavruša, M., Becot, J., Sperandio, O., & Miteva, M. A. et al. (2015). MTiOpenScreen: A web server for structure-based virtual screening. *Nucleic Acids Research*, *43*(W1), W448–W454. doi:10.1093/nar/gkv306 PMID:25855812

Lambert, C., Leonard, N., De Bolle, X., & Depiereux, E. (2002). ESyPred3D: Prediction of proteins 3D structures. *Bioinformatics (Oxford, England)*, *18*(9), 1250–1256. doi:10.1093/bioinformatics/18.9.1250 PMID:12217917

Macindoe, G., Mavridis, L., Venkatraman, V., Devignes, M.-D., & Ritchie, D. W. (2010). HexServer: An FFT-based protein docking server powered by graphics processors. *Nucleic Acids Research*, *38*(Web Serversuppl 2), W445–W449. doi:10.1093/nar/gkq311 PMID:20444869

Meng, X-Y., Zhang, H-X., Mezei, M, & Cui, M. (2011). Molecular Docking: A powerful approach for structure-based drug discovery. *Current Computational Aided Drug Design, 7*(2), 146–157.

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, *30*(16), 2785–2791. doi:10.1002/jcc.21256 PMID:19399780

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Morris, G. M., & Lim-Wilby, M. (2008). Molecular docking. *Methods in Molecular Biology (Clifton, N.J.)*, *443*, 365–382. doi:10.1007/978-1-59745-177-2_19 PMID:18446297

Nantasenamat, C., Isarankura-Na-Ayudhya, C., Naenna, T., & Prachayasittikul, V. (2009). A practical overview of quantitative structure-activity relationship. *EXCLI Journal*, *8*, 74–88.

Norel, R., Fischer, D., Wolfson, H. J., & Nussinov, R. (1994). Molecular surface recognition by a computer vision-based technique. *Protein Engineering*, *7*(1), 39–46. doi:10.1093/protein/7.1.39 PMID:8140093

Norel, R., Petrey, D., Wolfson, H. J., & Nussinov, R. (1999). Examination of shape complementarity in docking of unbound proteins. *Proteins*, *36*(3), 307–317. doi:10.1002/(SICI)1097-0134(19990815)36:3<307::AID-PROT5>3.0.CO;2-R PMID:10409824

O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., & Hutchison, G. R. (2011). Open Babel: An open chemical toolbox. *Journal of Cheminformatics*, *3*(1), 33. doi:10.1186/1758-2946-3-33 PMID:21982300

Oloff, S., Zhang, S., Sukumar, N., Breneman, C., & Tropsha, A. (2006). Chemometric analysis of ligand receptor complementarity: Identifying complementary ligands based on receptor information (COLIBRI). *Journal of Chemical Information and Modeling*, *46*(2), 844–851. doi:10.1021/ci050065r PMID:16563016

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605–1612. doi:10.1002/jcc.20084 PMID:15264254

Pierce, B. G., Wiehe, K., Hwang, H., Kim, B. H., Vreven, T., & Weng, Z. (2014). ZDOCK server: Interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics (Oxford, England)*, *30*(12), 1771–1773. doi:10.1093/bioinformatics/btu097 PMID:24532726

Rush, T. S., Grant, J. A., Mosyak, L., & Nicholls, A. (2005). A Shape-Based 3-D Scaffold Hopping Method and Its Application to a Bacterial Protein−Protein Interaction. *Journal of Medicinal Chemistry*, *48*(5), 1489–1495. doi:10.1021/jm040163o PMID:15743191

Schmidtke, P., Le Guilloux, V., Maupetit, J., & Tufféry, P. (2010). Fpocket: Online tools for protein ensemble pocket detection and tracking. *Nucleic Acids Research*, *38*(Web Server), W582–W589. doi:10.1093/nar/gkq383 PMID:20478829

Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., & Wolfson, H. J. (2005). PatchDock and SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Research*, *33*(Web Server), W363-367. doi:10.1093/nar/gki481 PMID:15980490

Schoch, G. A., Yano, J. K., Sansen, S., Dansette, P. M., Stout, C. D., & Johnson, E. F. (2008). Determinants of cytochrome p4502c8 substrate binding - structures of complexes with montelukast, troglitazone, felodipine, and 9-cis-retinoic acid. *The Journal of Biological Chemistry*, *283*(25), 17227–17237. doi:10.1074/jbc.M802180200 PMID:18413310

Schoch, G. A., Yano, J. K., Wester, M. R., Griffin, K. J., Stout, C. D., & Johnson, E. F. (2004). The Structure of Human Microsomal Cytochrome P450 3A4 Determined by X-ray Crystallography to 2.05-Å Resolution. *The Journal of Biological Chemistry*, *279*(37), 38091–38094. doi:10.1074/jbc.C400293200 PMID:15258162

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Schüttelkopf, A. W., & van Aalten, D. M. F. (2004). PRODRG: A tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallographica. Section D, Biological Crystallography*, *60*(8), 1355–1363. doi:10.1107/S0907444904011679 PMID:15272157

Shoichet, B. K., Kuntz, I. D., & Bodian, D. L. (2004). Molecular docking using shape descriptors. *Journal of Computational Chemistry*, *13*(3), 380–397. doi:10.1002/jcc.540130311

Söding, J. (2005). Protein homology detection by HMM-HMM comparison. *Bioinformatics (Oxford, England)*, *21*(7), 951–960. doi:10.1093/bioinformatics/bti125 PMID:15531603

Sun, H. (2008). Pharmacophore-based virtual screening. *Current Medicinal Chemistry*, *15*(10), 1018–1024. doi:10.2174/092986708784049630 PMID:18393859

Tan, K. P., Nguyen, B. T., Patel, S., Varadarajan, R., & Madhusudhan, M. S. (2013). Depth: A web server to compute depth, cavity sizes, detect potential small-molecule ligand-binding cavities and predict the pKa of ionizable residues in proteins. *Nucleic Acids Research*, *41*(W1), W314–W321. doi:10.1093/nar/gkt503 PMID:23766289

The UniProt Consortium. (2008). The Universal Protein Resource (UniProt). *Nucleic Acids Research*, *36*, D190–D195. PMID:18045787

Wallner, B., & Elofsson, A. (2005). All are not equal: A benchmark of different homology modeling programs. *Protein Science*, *14*(5), 1315–1327. doi:10.1110/ps.041253405 PMID:15840834

Wang, Y., Suzek, T., Zhang, J., Wang, J., He, S., Cheng, T., & Bryant, S. H. et al. (2014). PubChem BioAssay: Update. *Nucleic Acids Research*, *42*(1), D1075–D1082. doi:10.1093/nar/gkt978 PMID:24198245

Weininger, D. (1988). SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of Chemical Information and Modeling*, *28*(1), 31–63. doi:10.1021/ci00057a005

Willet, P., & Barnard, J. M. (1998). Chemical similarity searching. *Journal of Chemical Information and Computer Sciences*, *38*(6), 983–996. doi:10.1021/ci9800211

Wishart, D. S., Knox, C., Guo, A. C., Shrivastava, S., Hassanali, M., Stothard, P., & Woolsey, J. et al. (2006). DrugBank: A comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Research*, *34*(90001), D668–D672. doi:10.1093/nar/gkj067 PMID:16381955

Xu, D., Xu, Y., & Uberbacher, E. C. (2000). Computational tools for protein modeling. *Current Protein & Peptide Science*, *1*(1), 1–21. doi:10.2174/1389203003381469 PMID:12369918

Yu, J., Zhou, Y., Tanaka, I., & Yao, M. (2010). Roll: A new algorithm for the detection of protein pockets and cavities with a rolling probe sphere. *Bioinformatics (Oxford, England)*, *26*(1), 46–52. doi:10.1093/bioinformatics/btp599 PMID:19846440

This work was previously published in Methods and Algorithms for Molecular Docking-Based Drug Design and Discovery edited by Siavoush Dastmalchi, Maryam Hamzeh-Mivehroud, and Babak Sokouti, pages 360-379, copyright year 2016 by Medical Information Science Reference (an imprint of IGI Global).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 38
Mapping Population Health Management Roadmap into Cervical Cancer Screening Programs

Anastasius Moumtzoglou
Hellenic Society for Quality & Safety in Healthcare, Greece & P. & A. Kyriakou Children's Hospital, Greece

Abraham Pouliakis
ATTIKON University Hospital, Greece

ABSTRACT

Population Health Management (PHM) aims to provide better health outcomes for preventing diseases, closing care gaps and providing more personalized care. Since the inception of the Pap test, cervical cancer (CxCa) decreased in countries applying cervical cancer programs, involving both prevention and treatment. In this article, the authors map the PHM roadmap to the design of cervical cancer screening programs and examine the effect on the supporting information technology systems. Notwithstanding screening programs have a tight relation to PHM; the mapping reveals numerous interventions involving additional data sources, and timeless reconfiguration.

INTRODUCTION

'Public health' connotes a relatively narrow field with activities carried out by agencies granted with official functions. 'Population health' a term with broader content, is related to a field relevant to the study of several important factors for health. As a result, it involves many terms, such as outcomes, disparities, determinants, and risk factors (Kindig, 2007).

Notwithstanding the term 'population health' combines the concepts of both health and population, every term has an essential meaning of its own. The population is related to a group of persons being organized into numerous different units of analysis. Similarly, the term health was defined in a negative

DOI: 10.4018/978-1-5225-0549-5.ch038

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

manner, i.e. the absence of disease. Nowadays, the modern understanding stresses the positive aspects as well, and health is considered to be related to all life issues. Summarizing, today it is debated whether population health and public health are different or identical. Nevertheless, population health is defined as health outcomes and their distribution into a population (Kindig, 2007).

Population health management (PHM) has been defined as 'the technical field of endeavor which utilizes a variety of individual, organizational and cultural interventions to help improve the morbidity patterns (i.e., the illness and injury burden) and the health care use behavior of defined populations' (Hillman, 2002). It is differentiated from disease management because it includes (Howe & Spence, 2004):

- More chronic conditions and diseases.
- Uses a single point of contact and coordination and predictive modeling across multiple clinical conditions.

Moreover, PHM is considered a broader term than disease management, as it includes (Coughlin, Pope, & Leedle (Jr), 2006):

- Intensive care management for individuals at the highest level of risk.
- Personal health management for those at lower levels of predicted health risk.

At the provider level, Care Continuum Alliance (2012) highlights three components:

- The leadership and the central care delivery role of the primary care physician.
- The critical importance of patient activation.
- The capacity expansion of care coordination.

In this context, to successfully achieve all of these requirements, an organization should provide proactive, preventive and chronic care services to all managed patients. Additionally, this should take place both during encounters of patients with the healthcare system and in between. Therefore, providers should maintain regular contact with their patients and support them in the management of their health. Additionally, providers must manage patients at high risk, to prevent the deterioration of their health and avoid the development of complications. Finally, evidence-based protocols for the diagnosis and treatment of patients, in a consistent and cost-effective manner, are also required if for a provider-based PHM approach is followed.

Moreover, the Federal Agency for Healthcare Research and Quality (AHRQ) developed the concept of 'practice-based population health' (PBPH) and defined it as "an approach to care that uses information on a group of patients within a primary care practice or group of practices to improve the care and clinical outcomes of patients within that practice." (Cusack, Knudson, Kronstadt, Singer, & Brown, 2010).

To assist healthcare organizations, navigate the path toward the implementation of effective population health management solutions, Sanders (2014) suggested a roadmap of 12 steps:

- Organize accurate patient registries.
- Decide patient-provider acknowledgment.
- Define precise numerators in the patient registries.
- Follow and measure cost and clinical metrics.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

- Adhere to clinical practice guidelines.
- Engage in risk-management outreach.
- Obtain external data.
- Exchange information with patients.
- Educate and engage with patients.
- Develop and adhere to complex clinical practice guidelines.
- Coordinate between patient and care team.
- Track-specific outcomes.

Obviously, the implementation of health IT is among the essential components and requirements for planning and implementing PHM. The adoption of Electronic Health Records is just the first step toward the creation of the requisite infrastructure. A wide range of other IT applications to automate PHM, track results and engage the patients in their health care is required. Additionally, IT systems should be repeatedly evaluated because of: rapid technological changes, new government regulations and new approaches to patient management.

AHRQ recommends health IT tools for the stratification and monitoring of populations, as follows:

- Target patients being in greatest need of health services by stratifying the patients and narrowing subpopulations.
- Make patient related data actionable by generating alerts to the patients.
- Make data actionable by generating alerts towards health providers, about patient care needs.

Cervical cancer (CxCa) is the second most common cancer among women with age between 14 and 44 years. Additionally, it is the third cause of death related to cancer, after breast and lung cancer (Jemal et al., 2011). Infection by the Human papillomavirus (HPV) is strongly associated to CxCa. There are more than 100 HPV types. Thirteen are considered oncogenic and of high-risk for the development of CxCa (Malik, 2005; McCance, 1988; J. S. Smith et al., 2007). The virus is mainly sexually transmitted and affects millions of women worldwide. The disease is also very common in young population (approximately 70% of women will be infected during their lifetime). Nevertheless, this infection is usually transient and regresses. However, a high-risk HPV type and a compromised immune system constitute a fundamental factor for cervical intraepithelial neoplasia (CIN) or cancer (Riou et al., 1990; Wentzensen & Klug, 2008).

Moreover, many years are required for the development of precancerous lesions and progression to cancer (Ferenczy & Franco, 2002; Koshiol et al., 2008; Schlecht et al., 2001). Thus, there is adequate time for preventive actions before a neoplasia develops to cancer. Therefore, CxCa prevention seems to be an exciting field for the application of PHM methodologies.

Pap test is nowadays the most valuable tool for cervical cancer prevention (Brown, Bradbury, & Kraushaar, 1949; Hollenbeck, 1957; Hufford & Burns, 1948; Magendie, Cator, Partimbene, & Sentucq, 1955). It is performed by extracting painlessly cells from the cervix, and evaluating them on glass slides via a microscope. This assessment is a task that requires effort and can be accomplished by well-trained cytopathologists or cytotechnologists.

Due to the nature of CxCa, and the progressive evolution, screening programs can prevent cervical cancer (Syrjanen et al., 2010). However, this benefit can be achieved if only screening is well organized, including quality control and assurance procedures. Moreover, this can be effectively achieved if the

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

program is compliant with the PHM principles. The prevalence of cervical cancer, as a result of organized cervical cancer screening programs and the methods introduced by George Papanicolaou, decreased by 75% between 1955 and 1992. However, there are still many deaths for this reason. In Europe, each year more than 50,000 women develop CxCa, and about 25,000 die from this disease. The American Cancer Society data indicate that 12,200 new CxCa cases were expected in the United States in 2010 resulting in 4,210 deaths (American Cancer Society, 2010). In the UK, approximately 7-8% of the screened population have an abnormal smear (Paraskevaidis, Kyrgiou, & Martin-Hirsch, 2007; Statistical Bulletin: Cervical Screening Programme 2005-6, 2006). Out of these women, approximately 1.5-2% will present with a high-grade and 5% with a low-grade lesion. Women with high-grade lesions require medical treatment (surgical operation) and women with low-grade lesions require frequent monitoring because the lesion may regress or progress to high grade.

Obviously, CxCa screening is compliant with PHM concepts, as it involves prevention and treatment. Additionally, there are high-risk patients, for example, those with compromised immune system (immunosuppressed patients, patients with transplantations and patients with HIV).

The objective of the article is twofold:

- Present the concepts behind the design and implementation of cervical cancer screening programs, according to the European Guidelines (European Commission. Directorate-General for Health and Consumer Protection, 2008).
- Map them into the PHM principles.

MAPPING POPULATION HEALTH MANAGEMENT ROADMAP INTO CERVICAL CANCER SCREENING PROGRAMS

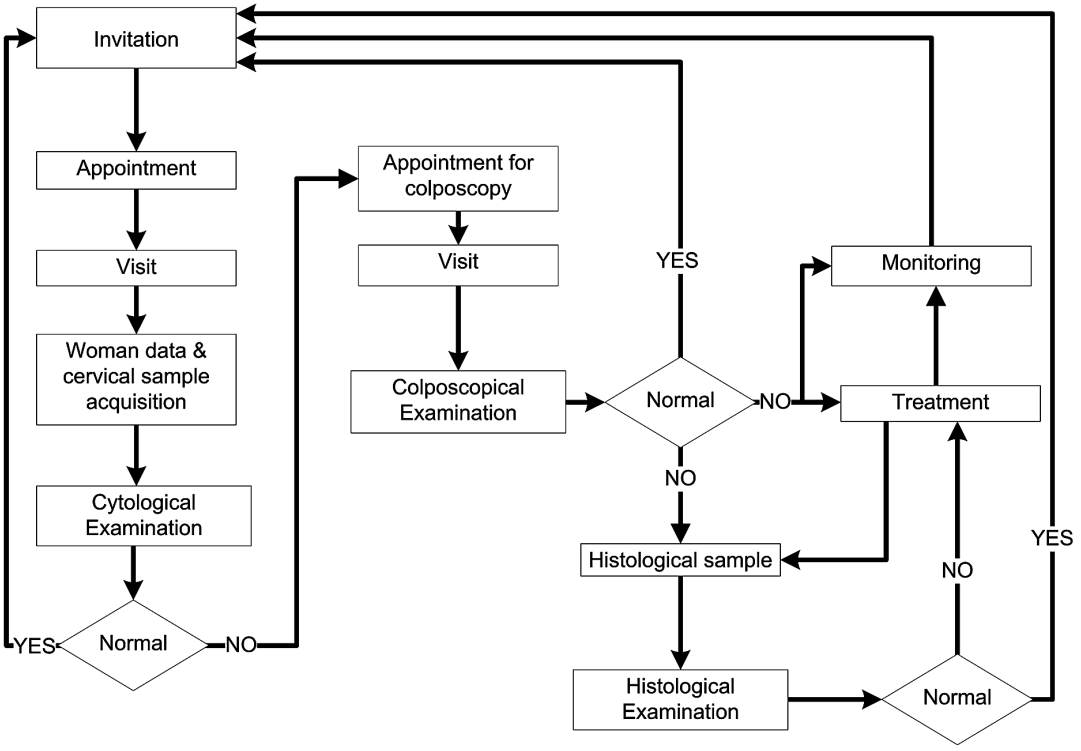
In this section, we analyze the potential application of the PHM roadmap (Sanders, 2014), on a point by point basis, in CxCa screening programs. The primary source of information is the European Guidelines for quality assurance in CxCa screening (European Commission. Directorate-General for Health and Consumer Protection, 2008). Many of the proposed procedures in the guidelines are compliant with the PHM roadmap. However, there are cases that gaps were found. In this case, we highlight them and propose solutions, based on PHM principles and IT solutions.

At an operational level, a CxCa screening program should be centralized. Ideally an IT system should serve the entire population of a country. The primary reason is that women under control may change their residence (Pouliakis, Karakitsou, Iliopoulou, Koutsouri, & Karakitsos, 2013). Web-based ICT systems can be used nationwide and may provide high security and adequate flexibility. However, continuous support is required if the deployment is for the first time. On the other side, a single point of control, as required by PHM, creates a burden in a single place. Solutions to overcome this limitation exist (Pouliakis & Karakitsos, 2012). For instance, hierarchical approaches may provide centralized control and simultaneously decentralized user training and support. That can reduce the workload of central offices and provide more personalized support, as required by the PHM principles.

The process of a CxCa screening program has a well-known and organized workflow (Figure 1), which can be formulated as follows (Athanasiadi et al., 2014; Pouliakis, Iliopoulou, & Karakitsou, 2012):

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 1. Workflow of the screening process



- During specific time intervals and depending on the woman cervical health status and age, each woman should be invited for the Pap test. This invitation is expected to motivate women to book a meeting to a health center and finally visit it for the Pap smear.
- During this visit, a specialist (usually a midwife or a gynecologist or eventually a nurse) records the woman's updated contact details, demographic data as well as her medical history. Additionally a cervical cytological sample is obtained and sent to a specialized cytopathology laboratory.
- In the cytopathology laboratory, the sample is prepared (staining) and subsequently is assigned to a specialized cytopathologist (in some countries nonmedical specialists – screeners are employed) for the Pap test. The result of this analysis is then sent back to the health center that the woman visited (or her family doctor).
- The result of the Pap examination is necessary for the subsequent management. According to this outcome, the time that the next invitation will be sent to the woman is decided. Moreover, it is determined if it is necessary to perform additional examinations. For example, to repeat the Pap test in a short period or to perform a colposcopic examination, sometimes including a biopsy. An abnormal result of the Pap smear poses different requirements for the woman management. Usually, the process points towards a review by a certified colposcopist (gynecological subspecialty) or to repeating the Pap test in a short time. In the first case, a new visit to the colposcopy clinic is arranged. If there is a normal result, the woman can be assigned to the usual cycle of invitations. However, in the case that the colposcopy result is abnormal a new sample, this time a tissue sample from the cervix (i.e. biopsy) is obtained and subsequently sent to a histology laboratory for further examination.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

- The outcome of the histological examination is then sent both to the health center that treats the woman and to the colposcopy clinic.
- If there are abnormal histological results, the colposcopy clinic is set accountable for the care of the woman (i.e. becomes the Accountable Care organization-ACO) and her treatment. If the histopathology result is within the normal limits, then the woman may return to the standard cycle and wait for the next invitation for the Pap test. In an abnormal result, a new visit to the colposcopy clinic for treatment should be arranged. After surgical treatment, tight monitoring for several years is required. In some cases, colposcopy is followed by immediate treatment (see and treat strategy).

This process is subsequently mapped to the 12 steps of the PHM roadmap.

Organize Accurate Patient Registries

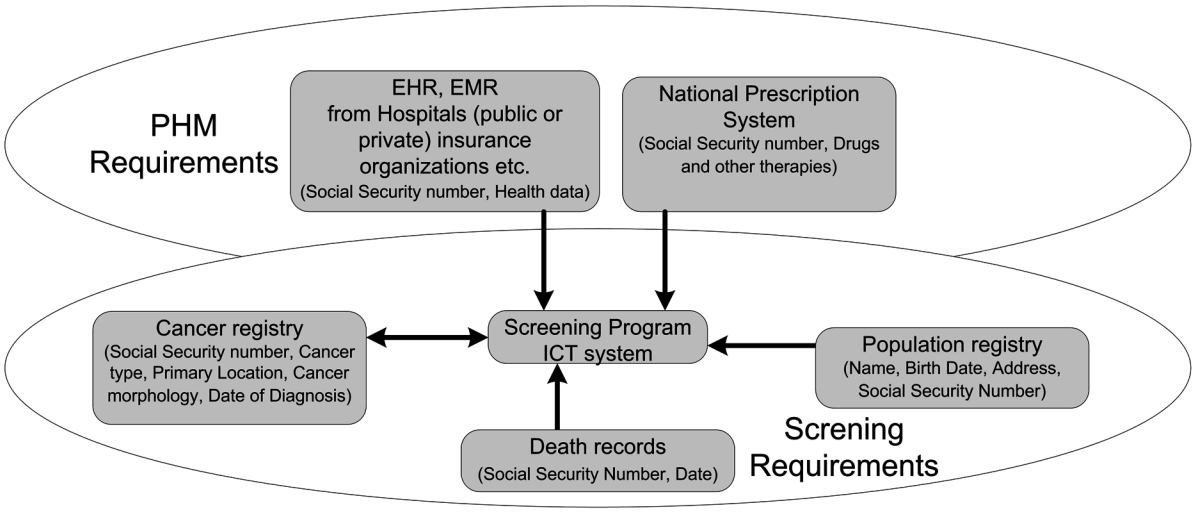
Definitely, each CxCa screening program starts with the definition of the population that should be under control (Athanasiadi et al., 2014; European Commission. Directorate-General for Health and Consumer Protection, 2008; Pouliakis et al., 2012). In fact, CxCa screening programs define, within the country, catchment areas. These are smaller geographical regions, used for finer control and precise management of the population and the screening project. The largest registry in this context is the population registry. Linkage of the IT system supporting the screening program to the national population registry is, therefore, an option. However, there are cases that such records are not always complete. A more practical solution could be the linkage with the social security number registry that provides a unique identification number for each person. This identification number is additionally/usually linked to the medical records of the women that, in fact, could be distributed in several hospitals around the country.

The ICT system has to be linked to various other registries as well (Figure 2). Depending on the availability of such systems, the preferred method is to connect them, at country level; otherwise such data should be stored and maintained within the screening program ICT system. As mentioned the primary interconnection is with the Population registry. Since the country population is dynamic and always changing, and the population ages evolve, it is more practical to obtain demographic data from external records that maintain those data, then store them within the ICT database. Therefore, the ICT system would be more efficient if it collects population information from databases relevant to new births, relocations and changes of communication details (if available). If population registries do not include death records, a linkage to appropriate registries should be maintained as well. A third major link is to cancer registry; thus, information related to cervical cancers (or cancers having as a primary source the cervix) can be retrieved, additionally the cancer registry can be automatically updated for new CxCa cases. Nowadays, there are registries related to neoplastic diseases as well; in this manner, High-grade Squamous Intraepithelial Lesion (HSIL) diagnoses may be forwarded to the relevant registries.

Technically, there are various methods available to communicate and exchange data. However, it is important to know which system initiates such data exchange; it is preferable that this communication be initiated immediately by the system that holds the new data or requests data from another system. For example, when a new cervical neoplastic case is found, it is the cancer registry that will initiate the communication and 'push' this information to the screening system. In contrast, if the screening system has the capability to store locally histological results, it is the screening system that will initiate this communication and feed the data to the cancer registry. Of course, data integrity should be performed

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 2. Interconnection of the Screening Program ICT system with other systems. Lower part: linkage requirements due to the screening program, upper part: additional requirements due to PHM.



automatically; in cases of deviations or compromised data, human operators should be alerted and take action.

However, PHM requirements for sharing of information relevant to patients' health problems, medications, laboratory results, and procedures, come to add another layer of information. Now it is necessary to have a more holistic approach; thus, it is not only the three links (population, death and cancer registries) required by CxCa screening programs that should be maintained. Additional links to obtain complete medical records, usually dispersed in numerous places, should be held as well (see Figure 2). For example, links with systems storing prescription data and hospitals or insurance organizations maintaining EHRs. This requirement poses additional difficulties because the increased sources of data increase the need for data integrity checks and maintenance. Thus, data providers have to improve data integrity by increasing the amount of data provided in the discrete form, and by the use of standards for data structures.

Decide Patient-Provider Acknowledgment

Identifying and defining the attributed providers, for each woman, initially seems to be a relatively easy process. The responsible team for every woman appears to be composed of (Figure 1):

- The midwife that performs a physical examination (macroscopic view) and extracts the cervical sample.
- The cytopathologist that performs the Pap test, and in abnormal cytological cases the gynecologist involved in colposcopy.

Moreover, in extreme situations, the histopathologist and a surgeon, usually the gynecologist that performed the colposcopy, are included on this team.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Algorithms determining the team members can be developed and put in practice by the use of the IT system supporting the CxCa screening process. Each health center has associations with one or more cooperating cytopathology laboratories as well as one or more colposcopy clinics or outpatient offices. That is a necessity to handle the cervical samples and additionally to instantly arrange meetings for women requiring gynecological consideration. These associations can be used to propose automatically and assign providers and doctors to manage women.

Additionally, PHM requirements, in the long term, require analysis of patient visit patterns to physicians, and moreover the evaluation of performance and financial indices. However, women participating in CxCa screening programs may have their own preferences to particular doctors (especially when visiting gynecologists). Additionally, an assignment to a particular doctor may be established or forced through the insurance organization. Obviously, the task of doctors handling a woman, even if she is participating in a screening program, may be out of the roster of physicians involved. Although the IT system may have the potential to propose and assign specific doctors, in reality women may visit other doctors. Thus, the analysis of visit patterns may face difficulties. The most difficult issue seems to be the collection of visits of women to individual physicians, in public or private hospitals, or even physician offices. The solution of this situation appears to be the proposed link to the prescription systems (Figure 2), where such data are stored. In some cases, a solution could be a legal framework, requiring recording of all visits and patient management actions to centralized systems for the population of a country.

Define Precise Numerators in the Patient Registries

According to the roadmap (Sanders, 2014), this requirement is pivotal in achieving physician involvement in patient management, as well as effectiveness. This condition is related to the capability of the IT system to identify large portions of the population having unique characteristics. There are numerous reasons for a patient to be included in particular groups and, therefore, unable to fully comply with medical protocols. These include:

- Language and communication difficulties (for example, emigrants or patients with cognitive problems).
- Physical inability.
- Economic issues.
- Refusal to participate despite knowing the underlying danger.
- Religion reasons.
- Medication contradictions.
- Geographic reasons.

European CxCa screening guidelines for quality assurance (European Commission. Directorate-General for Health and Consumer Protection, 2008) have anticipated such situations, and there is specific guidance to handle them:

- Women with disabilities should not be excluded from cervical screening because they are at equal risk to other women.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

- Assessment of women having individual needs.
- It is known that women who never had a sexual activity to be at low risk (as HPV is mainly sexually transmitted). However, because this is subject to change, all women should be invited irrespective of their sexual background.
- Hysterectomised women could be excluded from screening if surgery was not connected with cervical neoplasia. However, in case that the cervix is still in place, these women should be screened, as they may develop CxCa.
- Health professionals need to be sensitive and trained to handle the educational, linguistic and religious differences among women.

IT systems (Athanasiadi et al., 2014; Pouliakis, Athanasiadi, et al., 2014; Pouliakis, Margari, et al., 2014) and their implementation (Pouliakis et al., 2012) have already proposed the structures to support such situations. The relevant data can be supplied from participating women during their first visit, and stored in the database supporting the CxCa screening process. However, this is just only part of the information that can be available for women. There are numerous pieces of valuable information that should be provided by interconnected systems:

- Ethnicity can be provided from a population or social security number registries, and mortality from death records.
- Hysterectomised women from EHRs and women that had cervical cancer from cancer registries or registries of pre-neoplastic diseases.

The required information evolves with time as well as the guidelines change. Therefore, if the CxCa screening IT system is compliant with the PHM roadmap, it should incorporate a flexible architecture in many terms:

- Easy configuration to present new data.
- Linkage to new databases to retrieve information.
- Reconfigurable algorithms to identify the population to be included in particular groups or to be excluded.
- Production of alerts and forwarding to the appropriate health provider.
- Support for escalation of procedures in cases of deviations.

Furthermore, the IT process can be summarized as follows:

- Acquire data.
- Run algorithms to determine particular women and assign them to strata (flagging) relevant to their risk level.
- Identify the responsible clinician or assign one.
- Alert the woman and the physician.
- Follow the process according to specific patient management guidelines and in case of deviations alert.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Follow and Measure Cost and Clinical Metrics

This part of the roadmap is related to the metrics of the clinical outcomes and cost-effectiveness. According to the European Guidelines, the screening programs have numerous challenges; firstly, there are multiple actors involved: patients, health specialists with different disciplines, technical specialties and professionals from health and management authorities. Such programs are applied nationwide and affect the majority of the population. According to the European Guidelines, the information system is essential for managing the screening program and for automatic computation of numerous indicators. These can be ranked for effectiveness (Arbyn et al., 2007) of the screening method, as presented in Table 1.

Within the screening program, there are numerous indicators that are specified by the guidelines (European Commission. Directorate-General for Health and Consumer Protection, 2008). A design of an IT system supporting their implementation is also presented in the literature (Athanasiadi et al., 2014). Their grouping falls into three categories: Screening intensity, Screening tests performance and Diagnostic assessment and treatment. However, the prerequisites and costs of screening activities vary substantially between programs and are dependent on screening technologies. Additionally, a re-organization of activities and re-distribution of resources within the healthcare system may affect the final costs. For example, differences in sample taking (midwives vs. physicians), in sample pre-processing (conventional vs. liquid-based smear or HPV test), screening personnel (cytopathologists vs. screeners), the volume and distribution of tests, the excess consumption of tests and/or treatment, affect the cost-effectiveness (Hristova & Hakama, 1997).

In summary, the European guidelines provide a plethora of metrics for the clinical outcome. However, there are no defined metrics for cost effectiveness. The PHM roadmap, adds a financial layer besides the clinical layer. Thus, IT systems supporting the CxCa screening program should be adapted to support the required functionality. However, IT systems would face many difficulties in the analysis of cost metrics due to frequent changes. One solution would be to provide their data to other systems; that will calculate cost related metrics. Additionally, the models for the financial metrics must be continuously evaluated and adaptable to changes, due to advances in medical technology, changes in medical practices and government regulations. Thus, IT systems either for the screening program or the cost metrics should be adaptable to provide upgrades quickly.

Adhere to Clinical Practice Guidelines

PHM systems pose requirements on the management of each population cohort. Hopefully, CxCa screening programs have well-established procedures for managing women; these involve the age, the Pap

Table 1. Indicators of the effectiveness of cervical cancer screening methods (ranking)

1. Reduction of mortality (gained life-years)
2. Reduction of morbidity (incidence of cancer, quality-adjusted-life-years gained)
3. Reduction of incidence of cancer
4. Reduction of incidence of CIN3 or worst (CIN3+)
5. Increased detection rates for CIN2+ or CIN3+
6. Increased test positivity with increased, similar, or minor reduction of positive predictive value

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

test outcome, the colposcopy and histology results. The workflow described in Figure 1 is a relatively good mapping of the algorithms used and their influence on the design of IT systems. The majority of women that have a negative the Pap test will be screened after three to five years (according to the EU guidelines). Additionally, the age group starts at 25 or 30 years and ends to 60 or 65 years. Furthermore, there are guidelines from many national organizations (Ebina et al., 2015; Massad, 2003; R. A. Smith, Brooks, Cokkinides, Saslow, & Brawley, 2013) and numerous countries adopt them, especially for the management of abnormal cases.

The major issue concerning the IT systems is their adaptation to changes. The clinical practice guidelines vary from country to country; they are associated with the availability of financial and human resources; and quite often are revised (Vassilakos, Catarino, Frey Tirri, & Petignat, 2015). The second reason of changes is the evolution of new tests. For example, there are numerous studies evaluating HPV DNA or RNA testing (Coquillard, Palao, & Patterson, 2011; Duvlis et al., 2015; Monsonego et al., 2012). These tests aim to replace the Pap smear or drastically change the screening intervals as well as the patient management. Moreover, there are algorithms for patient triage based on the cumulative evaluation of many biomarkers (Bountris et al., 2014; Karakitsos et al., 2012; Pouliakis, Karakitsou, et al., 2014; Spathis et al., 2012). In all these situations, and irrelevant to specific reasons, IT systems should be easily modifiable. The best approach seems to be the use of pluggable algorithms; that evaluate the status of each woman and produce the appropriate recommendation for the next medical management step (or steps). Moreover, this approach would lead to personalized management.

Engage in Risk-Management Outreach

As mentioned earlier a PHM system should be capable of identifying and evaluating large portions of the population having distinctive characteristics (section: Define precise numerators in the patient registries). This functionality is relevant to the patients recorded in the CxCa screening IT system registry. However, this step of the roadmap goes beyond the registered patients; because over time, the data become richer, and new risk groups may be identified. The ultimate goal is to identify proactively individuals before they become members of the registry (i.e. patients).

CxCa screening is by design on this track. The goal is to detect women before they develop the disease and proactively act either by tight monitoring (as a lesion may regress) or by cure in the pre-neoplastic stage. On the other hand, the population involved in the CxCa IT system is enormous affecting all women in a country at the age of 20+. Additionally, the proposed data exchange with population registries augments this data set to all females from their birth date. Therefore, CxCa screening systems seem to engage women before they become patients and by concept to fulfill this requirement.

The major effort in the CxCa screening case in this aspect is the development of predictive modeling algorithms that forecast which patients are likely to have the disease and, therefore, result in significant health costs. Algorithms for identifying these cases even when there are missing data have been reported in the literature (Bountris et al., 2014; Karakitsos et al., 2012; Karakitsos et al., 2011; Kitchener et al., 2014; Komoto, Nishiwaki, Okamura, Tanaka, & Takebayashi, 2014; Pouliakis, Karakitsou, et al., 2014; Savage, Marlow, Alkis, & Simpson, 2014). Therefore, IT systems adapted to the PHM requirements should be capable of embedding such algorithms within a fast changing environment and lead towards their continuous evaluation in large datasets. Additionally automated alert mechanisms should be in place and management algorithms should be supported by the system along with their tight monitoring. The proposed concept of pluggable components in the IT system seems to be of interest in this case.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Obtain External Data

PHM roadmap suggests the usage of data being stored outside the hospitals managing the registered patients; due to a geography shift of healthcare from clinics to homes. CxCa screening is by its nature an extremely distributed process, engaging an organization body composed of health centers, cytopathology and histopathology laboratories and colposcopy/gynecology clinics distributed at a country level. In this case, the step should be translated as the acquisition of data outside of this organization body. Important data can be related to demographic information such as educational status, nutrition, and sexual lifestyle and habits, vaccination status or financial status. Some data cannot be acquired from external sources but should be provided by the women during their visits. However, a part of the data could be available from national registries especially those of educational and financial status. The primary barrier to such implementation is not technological. The business processes, the government regulations, the relationships among players and data holders, as well as agreements for data sharing makes this process more complicated at the managerial level than the technical. In any case, divergent technologies, such as Web services, database mediation systems, direct import from database tables, HL7 interconnectivity should be supported by the IT system in order to timely and easily link to external repositories.

Exchange Information with Patients

Although the proposed roadmap considers it a low ranked step, in CxCa screening it is fundamental. That happens because the screening process (in organized screening and not in opportunistic screening) is based on invitations (Figure 1). Therefore, it is crucial to have well-established means of communication. IT systems supporting this process can be relatively easily configured to generate messages and moreover dispatch them electronically to participating women. These methods include interfaces with e-mail systems (via well-established and secure protocols such as POP3 and SMTP and SMS dispatch systems). These messages can be sent to women according to their preferences on communication means. Another method involves the generation of printable invitation letters that can be sent via regular mail, especially for older women that are not familiar with digital technology. However, as the millennium population prevails, digital communication will be preferred over traditional means. An advantage of this approach is that modern e-mail systems and definitely SMS mechanisms have the capability of delivery confirmation. In contrast, regular mail not only has an associated time delay but requires registered letters, adding more cost to the already increased cost of the screening program.

PHM requires even more intensive communication with women; therefore, the use of voice and video communication becomes crucial. Thus, each woman's record should be coupled with extensive details for communication, enhanced with modern IP based communication tools. This any-time-any-place communication type increases patient engagement and can improve health outcomes. It is not the one-way communication (health actor - woman) that should be enhanced; communication initiated by women should be facilitated as well. Thus, call centers capable for all traditional and modern communication types should be in place to receive requests and route them to specialized personnel. Finally, a complete track of all communication activities is crucial to perform data analysis and improve this essential communication requirement.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Educate and Engage with Patients

Health related information is nowadays complicated due to the exponential growth of knowledge. This generates confusion and leads to difficulties in the process of decision-making (Arkin, 1999). As suggested in the previous section, the channels to provide information should be well-established. However, information provisioning to help women make decisions, requires new ways of interacting and communicating. An IT system supporting this process should support tailoring of information provided to women according to their profile; this can be dependent on their educational background, age, language barriers, religion and medical status. For example, the educational material targeted to a mid-aged woman with CIN1 and a young pregnant woman harboring CIN2 should be different. The solution for this PHM requirement comes through patient stratification into groups, and the assignment of the appropriate educational material for each case. A hierarchical model of material building could be of interest; as a population group can be further divided into smaller groups enabling a more personalized educational style. Use of multimedia educational material (according to the digital profile of the woman and preferences) is today preferable, because:

- Can be relatively cost efficient to produce.
- Can be distributed via the Internet with minimal cost.
- May educate women in a shorter time than traditional leaflets.
- It is available on the mobile device.

Moreover, multimedia material coupled with the GIS capabilities of mobile devices, is nowadays possible to contain URL links to arrange meetings at the nearest available medical center because the mobile device today is aware of the owner location. Therefore engagement of new women is facilitated. Concluding, the communication facilities coupled with the profile of each woman and the use of decision algorithms may be used to tailor educational material and have a significant effect on the CxCa screening process.

Develop and Adhere to Complex Clinical Practice Guidelines

CxCa screening has well-established guidelines, as it has evolved and matured since the fifties. Specifically there are:

- Established protocols for reporting cervical pathology (the Bethesda classification system - TBS2001 system) (Henry, 2003; Solomon & Nayar, 2004).
- A tiered classification system for reporting histological as well as colposcopical outcomes.
- Guidelines tailored to HPV infection status, vaccination, age, pregnancy and cervical medical history (Ebina et al., 2015; Massad, 2003; R. A. Smith et al., 2013).

As mentioned in section “Adhere to basic, clinical practice guidelines” both complex and basic these medical protocols and guidelines should be embeddable into the IT system. However, this PHM requirement increases complexity. One important factor is that HPV infection is usually transient. In more complex situations, for patients with co-morbidities (for example, immunosuppressed patients and women harboring HIV) there should be a rather different management approach. Nowadays, there are not

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

many clinical protocols that can account for patients with comorbidities. Usually, physicians are using their own approach or chain many protocols together. Obviously, guidelines will evolve as the medical literature becomes more focused (Ananworanich et al., 2015; Lambert et al., 2015; Ursu et al., 2015). In fact, IT systems supporting CxCa screening may have dual roles in this environment:

- To collect data for women, either via the medical history acquired during visits or from relevant registries.
- To alert physicians and enlist women in such a situation.

By way of these two functions, useful data can be collected and via their analysis new guidelines can be produced. Such guidelines should be 'pluggable' in the system in the form of new algorithms.

To the authors' opinion, CxCa screening programs optimized for comorbidities and the relevant complex protocols will have a comparative advantage and will differentiate themselves both in cost and clinical effectiveness.

Coordinate Between Patient and Care Team

Modern medicine requires that each patient be at the center of a plan supported by many specialists. In CxCa screening, the majority of the population is healthy women, which should be examined at regular intervals. In cases of deviations (i.e. non-compliance), these women should be alerted in an appropriate manner and motivated to return to the regular plan (an escalation procedure may be required). However, in cases of women harboring cervical lesions, the team is augmented and composed at least by two persons: the midwife, which has a close relationship with each woman and the colposcopist. In more severe situations, additional medical specialties can be involved (histopathologist and surgeon). PHM requires that this team be well-organized and all actors (including the woman) well-coordinated. Crucial to this step is the availability of communication tools, the completeness of the patient information, the convenient access to all data and eventually a tracking system that detects deviations from targets.

As the time factor is heavily involved and is crucial to the patient health, the IT system requires project management functionalities. If this concept is applied, each woman will participate in a separate project. This will have (as every other project): targets, tasks, milestones, risk management, and resources. Therefore, modern IT systems supporting screening seem to require embedding of project management functionalities.

Track Specific Outcomes

Data analysis is one of the primary requirements of PHM. It aims to measure and monitor specific results related to mortality, health status, prevalence, patient experience, and costs. All these reports are the cornerstone to measure efficiency and quality. The EU guidelines (European Commission. Directorate-General for Health and Consumer Protection, 2008) provide detailed tables to measure the project results mainly in the clinical aspect. Entities related to the implementation of such systems have been proposed (Athanasiadi et al., 2014) as well. PHM pushes CxCa screening programs one step further, because the requirements for such analyses, is not possible to be known in advance, i.e. during system design.

Fulfillment of this requirement seems to be out of the implementation of IT systems supporting CxCa screening. However, rule engines can be used to integrate the numerous data types required to perform

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

the analysis and OLAP (on-line analytical processing) techniques seem to be interesting in performing data mining and produce the required reporting. In that sense, the IT system should support random queries of stored information to extract data for further processing.

CONCLUSION

Population health management requires healthcare providers to develop new skills and new infrastructure for delivering care promising a holistic approach and more personalization. In this context, mapping the population health management roadmap into the cervical cancer screening programs reveals the essentiality of a radical redesign for the following reasons:

- The number of data sources in cervical cancer screening programs should increase in terms of their number and the complexity of data. However, the challenge is not technical as there are well-known and secure protocols to exchange information. The major issue is data integrity, automated translation in a standardized format and agreements between data owners.
- The IT systems should be easily reconfigurable to handle new data and incorporate new algorithms required at various levels. For example, protocols to apply and to identify patients in different risk groups.
- Data provisioning to external systems becomes more complicated, as reporting requirements and especially cost metrics is not included. Methods for self-describing data would be more efficient to fulfill this necessity.

Concluding, despite CxCa screening programs and the relevant information infrastructure, have by design and nature many principles required by PHM, there is a lack of support for many requirements. Adaptation of these systems, to be fully compliant to PHM needs, requires radical redesign and continuous support and evaluation by the involved stakeholders.

REFERENCES

American Cancer Society. (2010). Cancer Facts and Figures. Retrieved from http://www.cancer.org/acs/groups/content/@epidemiologysurveilance/documents/document/acspc-026238.pdf

Ananworanich, J., Prasitsuebsai, W., Kerr, S. J., Hansudewechakul, R., Teeratakulpisarn, N., Saisawat, K., & Phanuphak, N. et al. (2015). Cervical cytological abnormalities and HPV infection in perinatally HIV-infected adolescents. *J Virus Erad*, *1*(1), 30–37. PMID:26005716

Arbyn, M., Herbert, A., Schenck, U., Nieminen, P., Jordan, J., McGoogan, E., & Martin-Hirsch, P. et al. (2007). European guidelines for quality assurance in cervical cancer screening: Recommendations for collecting samples for conventional and liquid-based cytology. *Cytopathology*, *18*(3), 133–139. doi:10.1111/j.1365-2303.2007.00464.x PMID:17573762

Arkin, E. B. (1999). Cancer risk communication-what we know. *Journal of the National Cancer Institute. Monographs*, *1999*(25), 182–185. doi:10.1093/oxfordjournals.jncimonographs.a024200 PMID:10854479

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Athanasiadi, E., Pouliakis, A., Bilirakis, E., Fragoulakis, V., Bilirakis, C., & Karakitsos, P. (2014). Quality Indicators for Cervical Cancer Screening Programs: The Role of Computerized Systems.[IJRQEH]. *International Journal of Reliable and Quality E-Healthcare*, *3*(4), 19–37. doi:10.4018/ijrqeh.2014100102

Bountris, P., Haritou, M., Pouliakis, A., Margari, N., Kyrgiou, M., Spathis, A., & Koutsouris, D. D. et al. (2014). An intelligent clinical decision support system for patient-specific predictions to improve cervical intraepithelial neoplasia detection. *Biomed Res Int*, *341483*. doi:10.1155/2014/341483 PMID:24812614

Brown, W. E., Bradbury, J. T., & Kraushaar, O. F. (1949). The Papanicolaou test in the cancer control program. *The Journal of the Kansas Medical Society*, *50*(5), 217–221. PMID:18120901

Care Continuum Alliance. (2012). Advancing the Population Health Improvement Model. Retrieved from http://www.fiercehealthit.com/story/hennepin-health-project-looks-build-countywide-ehr-program-national-implica/2012-01-10

Coquillard, G., Palao, B., & Patterson, B. K. (2011). Quantification of intracellular HPV E6/E7 mRNA expression increases the specificity and positive predictive value of cervical cancer screening compared to HPV DNA. *Gynecologic Oncology*, *120*(1), 89–93. doi:10.1016/j.ygyno.2010.09.013 PMID:20950847

Coughlin, J. F., Pope, J. E., & Leedle, B. R. Jr. (2006). Old Age, New Technology, and Future Innovations in Disease Management and Home Health Care. *Home Health Care Management & Practice*, *18*(3), 196–207. doi:10.1177/1084822305281955

Cusack, C. M., Knudson, A. D., Kronstadt, J. L., Singer, R. F., & Brown, A. L. (2010). *Practice-Based Population Health: Information Technology to Support Transformation to Proactive Primary Care* (p. 44). Retrieved from https://pcmh.ahrq.gov/sites/default/files/attachments/Information%20Technology%20to%20Support%20Transformation%20to%20Proactive%20Primary%20Care.pdf

Duvlis, S., Popovska-Jankovic, K., Arsova, Z. S., Memeti, S., Popeska, Z., & Plaseska-Karanfilska, D. (2015). HPV E6/E7 mRNA versus HPV DNA biomarker in cervical cancer screening of a group of Macedonian women. *Journal of Medical Virology*, *87*(9), 1578–1586. doi:10.1002/jmv.24199 PMID:25880030

Ebina, Y., Yaegashi, N., Katabuchi, H., Nagase, S., Udagawa, Y., Hachisuga, T., & Yoshikawa, H. et al. (2015). Japan Society of Gynecologic Oncology guidelines 2011 for the treatment of uterine cervical cancer. *International Journal of Clinical Oncology*, *20*(2), 240–248. doi:10.1007/s10147-015-0806-7 PMID:25800808

European Commission. Directorate-General for Health and Consumer Protection. (2008). M. Arbyn, A. Anttila, J. Jordan, G. Ronco, U. Schenck, N. Segnan, H. G. Wiener, A. Herbert, J. Daniel & L. von-Karsa (Series Eds.), *European guidelines for quality assurance in cervical cancer screening* (pp. 291). Retrieved from http://bookshop.europa.eu/en/european-guidelines-for-quality-assurance-in-cervical-cancer-screening-pbND7007117/

Ferenczy, A., & Franco, E. (2002). Persistent human papillomavirus infection and cervical neoplasia. *The Lancet Oncology*, *3*(1), 11–16. doi:10.1016/S1470-2045(01)00617-9 PMID:11905599

Henry, M. R. (2003). The Bethesda System 2001: An update of new terminology for gynecologic cytology. *Clinics in Laboratory Medicine*, *23*(3), 585–603. doi:10.1016/S0272-2712(03)00053-2 PMID:14560530

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Hillman, M. (2002). Testimony before the Subcommittee on Health of the House Committee on Ways and Means, Hearing on Promoting Disease Management in Medicare. Retrieved from http://waysandmeans.house.gov/Legacy/health/107cong/4-16-02/4-16hill.htm

Hollenbeck, Z. J. (1957). The value of the Papanicolaou smear test. *Henry Ford Hospital Medical Bulletin*, *5*(4), 239–242. PMID:13501756

Howe, R., & Spence, C. (2004). Population Health Management: Healthways PopWorks (Vol. 2, pp. 291–297). HCT Project; Retrieved from http://mthink.com/article/population-health-management-healthways-popworks/

Hristova, L., & Hakama, M. (1997). Effect of screening for cancer in the Nordic countries on deaths, cost and quality of life up to the year 2017. *Acta Oncologica (Stockholm, Sweden)*, *36*(Suppl. 9), 1–60. PMID:9143316

Hufford, C. E., & Burns, E. L. (1948). Papanicolaou test in the early diagnosis of uterine cancer. *Ohio Medicine*, *44*(9), 900–903. PMID:18879503

Jemal, A., Bray, F., Center, M. M., Ferlay, J., Ward, E., & Forman, D. (2011). Global cancer statistics. *CA: a Cancer Journal for Clinicians*, *61*(2), 69–90. doi:10.3322/caac.20107 PMID:21296855

Karakitsos, P., Chrelias, C., Pouliakis, A., Koliopoulos, G., Spathis, A., Kyrgiou, M., & Paraskevaidis, E. et al. (2012). Identification of women for referral to colposcopy by neural networks: A preliminary study based on LBC and molecular biomarkers. *Journal of Biomedicine & Biotechnology*, *303192*. doi:10.1155/2012/303192 PMID:23093840

Karakitsos, P., Pouliakis, A., Meristoudis, C., Margari, N., Kassanos, D., Kyrgiou, M., & Paraskevaidis, E. et al. (2011). A preliminary study of the potential of tree classifiers in triage of high-grade squamous intraepithelial lesions. *Analytical and Quantitative Cytology and Histology*, *33*(3), 132–140. PMID:21980616

Kindig, D. A. (2007). Understanding population health terminology. *The Milbank Quarterly*, *85*(1), 139–161. doi:10.1111/j.1468-0009.2007.00479.x PMID:17319809

Kitchener, C. H., Canfell, K., Gilham, C., Sargent, A., Roberts, C., Desai, M., & Peto, J. (2014). The clinical effectiveness and cost-effectiveness of primary human papillomavirus cervical screening in England: Extended follow-up of the ARTISTIC randomised trial cohort through three screening rounds. *Health Technology Assessment*, *18*(23), 1–196. doi:10.3310/hta18230 PMID:24762804

Komoto, S., Nishiwaki, Y., Okamura, T., Tanaka, H., & Takebayashi, T. (2014). Effect size and cost-effectiveness estimates of breast and cervical cancer screening reminders by population size through complete enumeration of Japanese local municipalities. *BMC Public Health*, *14*(1), 43. doi:10.1186/1471-2458-14-43 PMID:24433346

Koshiol, J., Lindsay, L., Pimenta, J. M., Poole, C., Jenkins, D., & Smith, J. S. (2008). Persistent human papillomavirus infection and cervical neoplasia: A systematic review and meta-analysis. *American Journal of Epidemiology*, *168*(2), 123–137. doi:10.1093/aje/kwn036 PMID:18483125

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Lambert, C. C., Chandler, R., McMillan, S., Kromrey, J., Johnson-Mallard, V., & Kurtyka, D. (2015). Pap test adherence, cervical cancer perceptions, and HPV knowledge among HIV-infected women in a community health setting. *The Journal of the Association of Nurses in AIDS Care*, *26*(3), 271–280. doi:10.1016/j.jana.2014.11.007 PMID:25769754

Magendie, J., Cator, R., Partimbene, M., & Sentucq, M. (1955). [Experience with Papanicolaou test in diagnosis of female genital cancers]. *Comptes Rendus de la Société Francaise de Gynecologie*, *25*(6), 264–271. PMID:13285014

Malik, A. I. (2005). The role of human papilloma virus (HPV) in the aetiology of cervical cancer. *JPMA. The Journal of the Pakistan Medical Association*, *55*(12), 553–558. PMID:16438278

Massad, L. S. (2003). Guidelines for cervical cancer screening from the American cancer society: The Earth is moving. *Journal of Lower Genital Tract Disease*, *7*(2), 87–88. doi:10.1097/00128360-200304000-00002 PMID:17051050

McCance, D. J. (1988). Human papillomavirus (HPV) infections in the aetiology of cervical cancer. *Cancer Surveys*, *7*(3), 499–506. PMID:2854000

Monsonego, J., Hudgens, M. G., Zerat, L., Zerat, J. C., Syrjanen, K., & Smith, J. S. (2012). Risk assessment and clinical impact of liquid-based cytology, oncogenic human papillomavirus (HPV) DNA and mRNA testing in primary cervical cancer screening (the FASE study). *Gynecologic Oncology*, *125*(1), 175–180. doi:10.1016/j.ygyno.2012.01.002 PMID:22233689

Paraskevaidis, E., Kyrgiou, M., & Martin-Hirsch, P. (2007). Have we dismissed ablative treatment too soon in colposcopy practice? *BJOG*, *114*(1), 3–4. doi:10.1111/j.1471-0528.2006.01178.x PMID:17233854

Pouliakis, A., Athanasiadi, E., Karakitsou, E., Archondakis, S., Mourtzikou, A., Stamouli, M., & Karakitsos, P. et al. (2014). ISO 15189:2012 Management Requirements for Cytopathology Laboratory Information Systems. *International Journal of Reliable and Quality E-Healthcare*, *3*(3), 37–57. doi:10.4018/ijrqeh.2014070103

Pouliakis, A., Iliopoulou, D., & Karakitsou, E. (2012). Design and Implementation Issues of an Information System Supporting Cervical Cancer Screening Programs. *Journal of Applied Medical Sciences*, *1*(1), 61–91.

Pouliakis, A., & Karakitsos, P. (2012, 30 September - 3 October 2012). *Advantages Of A Hierarchical Software Deployment Strategy For The Hellenic Cervical Cancer Prevention Screening Program.* Paper presented at the 37th European Congress of Cytology, Dubrovnik-Cavtat, Croatia.

Pouliakis, A., Karakitsou, E., Chrelias, C., Pappas, A., Panayiotides, I., Valasoulis, G., & Karakitsos, P. et al. (2014). The Application of Classification and Regression Trees for the Triage of Women for Referral to Colposcopy and the Estimation of Risk for Cervical Intraepithelial Neoplasia: A Study Based on 1625 Cases with Incomplete Data from Molecular Tests. *Biomed Res Int*, *2014*, 914740. PMID:26339651

Pouliakis, A., Karakitsou, E., Iliopoulou, D., Koutsouri, G., & Karakitsos, P. (2013, May 26–30). *Roaming Women In The Hellenic Population Based Cervical Cancer Screening Program. Secure Access Of Medical Records And High Flexibility.* Paper presented at the 18th International Congress of Cytology, Paris, France.

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Pouliakis, A., Margari, N., Spathis, A., Kottaridi, C., Stamouli, M., Mourtzikou, A., & Karakitsos, P. et al. (2014). ISO 15189:2012 Technical Requirements for Cytopathology Laboratory Information Systems. *International Journal of Reliable and Quality E-Healthcare*, *3*(3), 58–80. doi:10.4018/ijrqeh.2014070104

Riou, G., Favre, M., Jeannel, D., Bourhis, J., Le Doussal, V., & Orth, G. (1990). Association between poor prognosis in early-stage invasive cervical carcinomas and non-detection of HPV DNA. *Lancet, 335*(8699), 1171-1174.

Sanders, D. (2014). A Landmark, 12-Point Review of Population Health Management Companies (Executive Report). 1-17. Retrieved from https://www.healthcatalyst.com/landmark-review-of-population-health-management-companies

Savage, A. H., Marlow, N. M., Alkis, M. H., & Simpson, K. N. (2014). Cost-effectiveness analysis of 2 surveillance options for cervical intraepithelial neoplasia 1. *Journal of Lower Genital Tract Disease*, *18*(2), 136–141. doi:10.1097/LGT.0b013e31829eaf31 PMID:24270194

Schlecht, N. F., Kulaga, S., Robitaille, J., Ferreira, S., Santos, M., Miyamura, R. A., & Franco, E. L. et al. (2001). Persistent human papillomavirus infection as a predictor of cervical intraepithelial neoplasia. *Journal of the American Medical Association*, *286*(24), 3106–3114. doi:10.1001/jama.286.24.3106 PMID:11754676

Smith, J. S., Lindsay, L., Hoots, B., Keys, J., Franceschi, S., Winer, R., & Clifford, G. M. (2007). Human papillomavirus type distribution in invasive cervical cancer and high-grade cervical lesions: A meta-analysis update. *International Journal of Cancer*, *121*(3), 621–632. doi:10.1002/ijc.22527 PMID:17405118

Smith, R. A., Brooks, D., Cokkinides, V., Saslow, D., & Brawley, O. W. (2013). Cancer screening in the United States, 2013: A review of current American Cancer Society guidelines, current issues in cancer screening, and new guidance on cervical cancer screening and lung cancer screening. *CA: a Cancer Journal for Clinicians*, *63*(2), 88–105. doi:10.3322/caac.21174 PMID:23378235

Solomon, D., & Nayar, R. (2004). *The Bethesda system for reporting cervical cytology: definitions, criteria, and explanatory notes* (2nd ed.). New York: Springer. doi:10.1007/978-1-4612-2042-8

Spathis, A., Kottaridi, C., Chranioti, A., Meristoudis, C., Chrelias, C., Panayiotides, I. G., & Karakitsos, P. et al. (2012). mRNA and DNA detection of human papillomaviruses in women of all ages attending two colposcopy clinics. *PLoS ONE*, *7*(11), e49205. doi:10.1371/journal.pone.0049205 PMID:23166611

. *Statistical Bulletin: Cervical Screening Programme 2005-6*. (2006). England.

Syrjanen, K., Di Bonito, L., Goncalves, L., Murjal, L., Santamaria, M., Mahovlic, V., & Schmitt, F. C. et al. (2010). Cervical cancer screening in Mediterranean countries: Implications for the future. *Cytopathology*, *21*(6), 359–367. doi:10.1111/j.1365-2303.2010.00795.x PMID:20718841

Ursu, R. G., Onofriescu, M., Luca, A., Prisecariu, L. J., Salceanu, S. O., Nemescu, D., & Iancu, L. S. (2015). The Need for Cervical Cancer Control in HIV-Positive and HIV-Negative Women from Romania by Primary Prevention and by Early Detection Using Clinically Validated HPV/DNA Tests. *PLoS ONE*, *10*(7), e0132271. doi:10.1371/journal.pone.0132271 PMID:26186361

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Vassilakos, P., Catarino, R., Frey Tirri, B., & Petignat, P. (2015). Cervical cancer screening in Switzerland: Time to rethink the guidelines. *Swiss Medical Weekly*, *145*, w14112. doi:10.4414/smw.2015.14112 PMID:25664576

Wentzensen, N., & Klug, S. J. (2008). Early detection of cervical carcinomas: Finding an overall approach. *Deutsches Ärzteblatt International*, *105*(37), 617–622. doi:10.3238/arztebl.2008.0617 PMID:19471627

This work was previously published in the International Journal of Reliable and Quality E-Healthcare (IJRQEH), 4(3); edited by Anastasius Moumtzoglou, pages 1-18, copyright year 2015 by IGI Publishing (an imprint of IGI Global).

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Index

3D-QSAR 37, 40-43, 45-46, 64, 91-92, 94, 136, 146, 148, 424-425, 437, 441, 491-493, 500-501, 504, 511, 608, 615, 618-619, 782
6-Fluoroquinolones 445, 453, 480

A

accelrys 133-134, 438, 444, 576, 896
AChE enzyme 597, 599, 604, 616, 624, 636
actinic 557
activism 638, 645, 739, 742-744, 750-751
Advanced Glycation End-Products (AGE) 235, 265
affine transformation 698-699, 708
age-related macular degeneration 366, 372, 380
antibacterial agents 46, 480
antibody-drug conjugates 346, 350
antimicrobial 91, 324
antioxidants 241, 316, 408-412, 416, 418-419, 421-428, 432, 592
aromatase 118-121, 124-134, 136-139, 141-156, 158-163, 186
Artificial Neural Networks (ANN) 22, 36, 89, 409, 437, 442, 457-458, 480, 563, 575, 582
ATI-1123 388, 407
autonomy 203, 206, 641, 649, 729-730, 738, 743
avatar 676-687
avatar creation 676-680, 687

B

benchmark 834, 921, 923-924, 928, 939
beneficence 728-729, 738
binding affinity 94, 118, 135, 372, 420-421, 451, 460, 499, 506-509, 576, 578, 618-619, 781, 803-804, 812, 814, 829, 831-834, 836, 840-841, 848, 860-861, 865, 878, 893, 897, 915-917, 920-923, 927-929
binding energy 487, 577, 619, 812, 850, 854, 859, 861, 867, 878, 880, 882, 925
binding free energy 578, 829, 831, 834, 836-837, 840-842, 848, 867, 880, 916, 918-920, 924-925, 927, 952-953
binding pose 451, 812, 817, 856, 915-916, 922, 927-928
binding specificity 776, 917, 924-925, 928-929
biomedical engineering 311, 352-363, 365
biopsy 123, 268, 275, 557
bioremediation 313, 340, 817
breast cancer 33, 118-119, 121-127, 131-132, 162-163, 186, 228, 234, 242, 346, 370, 373, 388-389, 391-393, 637-640, 643-650, 662-664, 667, 692-693, 704, 739-751, 884
breast cancer screening 637, 639-640, 643-644, 650, 740, 742, 745-747

C

Carbon Nanotubes (CNT) 95, 326, 340, 342, 354, 386, 891, 902
cardiovascular 204, 217-218, 220, 222, 229, 231-232, 236, 238, 241, 245, 266, 689
CASF 924
CDP 384, 407
cervical cancer screening 276, 283, 960, 963, 974
chemoinformatics 593, 614
CIN 267-268, 270, 273, 278, 283, 285, 287, 663, 962, 972
closed ICU 718, 727
cloud computing 187-189, 197-199, 576, 582
collaborative care delivery 187-189, 192, 197, 199
CoMFA 35, 41, 43, 92-93, 97, 118, 135, 137, 142, 145-147, 159-160, 162, 186, 425, 437, 441, 491-492, 499-501, 503-504, 575, 608, 618-620
communicable diseases 662, 666

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

community based care delivery 187-189
competitive inhibitors 120, 132, 186, 507, 595, 599
complementarity 30, 415, 577, 597, 765, 771, 834, 836, 862, 876-877, 879, 883, 889, 928, 946-947, 949-950
Computer Aided-Drug Design (CADD) 409, 412, 829, 831, 877, 883
CoMSIA 35, 41, 92-93, 118, 137, 142-143, 147, 160, 162, 186, 424, 441, 491-492, 499-501, 503-504, 575, 608, 618-620
conformational entropy 899, 918, 927
conformational space 775-776, 808, 834, 855-856, 876, 879-880, 890, 896, 901, 906, 939, 947
consultative model 2, 8, 18
contrast enhancement 547
correlation coefficient 35, 38, 40, 88, 414, 418, 421, 424, 427, 442, 502-503, 531, 533, 612-613, 617, 619-620, 698, 814, 923
coumarin derivatives 136, 409, 418, 426-428
critical care billing 709, 719
critical care team 709, 711, 727
CSAR 141, 924
curcumin 409, 418, 424-425, 428, 433
CXCA 267-269, 287, 289, 960, 962-963, 965-974

D

data analysis 21-22, 38, 40, 43, 91, 289, 414, 422, 571, 579-580, 590, 684, 971, 973
Daubechies Complex Wavelet Transform (DCWT) 519, 522, 524, 527, 532-533, 536
decision-making 2, 6-8, 19, 203, 207, 641-643, 728, 730, 732-733, 736, 742, 746, 748, 972
deka 892, 902, 904
dermatology 544, 554, 557
dermoscopic 542, 545, 548, 550, 554, 559, 562-563
dermoscopic images 542, 545, 554, 559, 563
descriptors 20-23, 26, 28-37, 40-43, 45-48, 64, 78, 86-93, 95, 141, 149, 151, 409, 413-416, 418-428, 433, 437, 441-443, 457, 484, 487, 489-490, 501-506, 518, 575, 607-610, 612, 616, 619-620, 622-624, 814, 861, 883, 903-905, 921-922, 947
desired self 676-678, 680, 684
Diabetes 204, 216-218, 220, 229-233, 235, 238-241, 243-245, 346, 373, 381, 638, 661, 690, 830
Diabetic Retinopathy 366, 373, 381, 554
diagnosis 3, 119, 194, 207, 280, 283, 285, 289, 302-303, 305, 324, 519-523, 536-537, 542-546, 550-551, 554-555, 557, 560, 562, 581, 640, 644, 660, 665, 669, 690, 718, 740-742, 746, 961
dimensionality reduction 64
Discrete Cosine Transform (DCT) 522, 524, 532-533, 536-537
Discrete Fourier Transform (DFT) 522, 524, 532-533, 536-537
Discrete Wavelet Transform (DWT) 521-522, 524-526, 532-533, 536-537
DNA Gyrase 434, 436, 447, 453-456, 459-463, 465, 480
DNA nanotechnology 342, 348-350
docking 21, 26, 38, 44, 94-95, 118, 135, 138-142, 144-145, 148, 154, 158-159, 162-163, 186, 428, 437-438, 448-449, 451-452, 459-460, 462-463, 484, 486-487, 491-496, 498-499, 501-503, 510-512, 518, 571, 573, 576-579, 590, 593, 597, 606, 615, 618-620, 623, 764-766, 771-779, 781-783, 803-806, 808-815, 817, 819, 829, 831-837, 840-842, 848-869, 876-886, 889, 891-893, 895-906, 915-918, 922, 924, 926-929, 939, 941-942, 945-950, 952-954
drug delivery 311-312, 324, 326-327, 344, 349, 353, 355-356, 358, 363-364, 366-367, 373-374, 382-384, 386-394, 891, 902-903
drug design 20-21, 37, 40, 46-48, 67, 78, 93, 97-98, 118, 121-122, 133, 135, 186, 409, 412-413, 428, 434, 436-439, 451, 456, 464, 484, 487, 497-498, 506-507, 511-512, 571, 575, 577-578, 591, 593, 608, 614-615, 620, 624, 764-766, 772, 781-783, 803, 813, 829, 831, 836, 848-849, 851, 862, 865, 877-878, 883, 886, 891, 893, 899, 901-902, 915-917, 929, 942, 944-946, 954
drug discovery 20-21, 94, 121, 367, 412-413, 434-438, 440, 444, 446, 448, 464-465, 480, 482, 485, 512, 574, 579, 607, 614, 765, 777, 781, 783, 804, 819, 829-831, 833, 848, 850, 868, 876, 878, 883, 885-886, 891-893, 895, 902, 905, 915, 926, 928-929, 941-943, 955
drug-likeness 20, 47, 437, 446, 457, 480
Dual Tree Complex Wavelet Transform (DT-CWT) 522, 524, 526-527, 532-533, 536
Dysplastic Nevi 550, 554, 557

E

early detection 123, 367, 543, 550, 563, 566, 639-640, 646-649, 660-661, 665-669, 740-741, 744, 747, 749
ECM 369, 394, 407
electronic portal imaging 693, 708
empowerment 668, 739, 741-743, 750
end-of-life care 1-7, 9, 18, 733
end of life choice 203-205, 214-215

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

environmental sciences 82-83, 97, 324, 608
EPR 385-387, 392-396, 407
ethics 4, 637-638, 640-642, 648, 728-729, 733, 736

F

feature extraction 284, 303, 423, 519, 529, 542-543, 545, 548-549, 555, 561-562, 564
feature selection 32, 47-48, 414, 423, 543, 550-551, 555, 619
flavonoids 141-143, 161, 409-411, 416, 418-422, 428, 433
FlexE 858, 901
force field 41, 142, 451, 486-487, 491, 578, 616, 779-780, 811-812, 834, 837, 859, 876, 880, 882, 889, 901
fragment docking 804, 848-849, 865, 899
free radicals 232, 409-410, 422-423, 426, 433, 592
fuzzy C mean 302, 304

G

GAGs 407
generalist palliative care 2, 9, 18
genomics 366, 381, 571, 590, 849
GFA 426, 501, 503, 505, 609, 612, 620-623, 636
Glaucoma 366, 368-370, 381

H

Haem Iron 217, 222, 229, 232-235, 265-266
HCV inhibitors 512, 518
healthcare 187-189, 199, 203-204, 210, 311, 352, 362, 382, 483, 637-639, 641-643, 646-650, 660-661, 664-669, 710, 713, 729, 739-751, 961, 969, 971, 974
health care agent 731, 738
health communication 637, 642, 645-646, 747-748
Health Information Technology (HIT) 187-188
heath care proxy 738
Hepatitis C Virus 482-483, 491, 518, 662, 782, 818
Heterocyclic Amines (HCA) 229, 232-233, 266
high intensity staffing model 720, 727
High-Reliability Organization (HRO) 727
histogram 149, 302-306, 308, 547
homology modeling 145, 162, 448-449, 480, 590, 615, 804, 832, 865, 867-868, 946
homology models 450, 459-462, 848-849, 864-865, 868
human organ trafficking 215
human trafficking 203-204, 215
hydroxybenzalacetones 409, 418, 425, 428

I

image registration 520, 522-523, 689, 692, 694-695, 708
induced-fit concept 802
in silico 33, 45, 48, 80, 85, 135, 152, 426, 428, 433-440, 444, 447-448, 450, 452-453, 456-457, 462, 464-465, 480, 482, 485, 492, 495, 506, 512, 591, 593, 614, 636, 765, 781, 783, 829, 834, 836, 867-868, 884, 886, 902, 955
in silico methods 80, 428, 433-434, 452, 593
Insulin 230-233, 236, 240, 242
integrated in silico methods 434
integrative model 2, 8, 18
isocenter 689, 691-692, 694, 697, 699-705, 708

L

lead identification 162, 623, 782-783, 890
lesion 267-268, 280-281, 283, 285, 287, 289, 302-303, 306-308, 542-546, 548-549, 552, 554-555, 558-562, 963, 965, 970
life-sustaining therapy
  withdrawal 8, 732, 736, 738
  withholding 731-732, 738
ligand 21, 44, 94-95, 118, 121, 134-135, 141, 148, 157, 186, 383-384, 386-387, 396, 412-413, 415, 437, 439, 448-452, 459-463, 465, 485-487, 490, 492, 496, 502-503, 505-506, 518, 573-579, 590, 596, 615-616, 618, 765-766, 771-772, 774-775, 778, 780-781, 783, 802-805, 808-814, 829, 831-837, 839-842, 848-861, 865, 867-868, 876-880, 882-886, 889-893, 895-903, 905-906, 915-918, 920-922, 924-925, 927-929, 939-942, 945-947, 949-950, 952-954
ligand based drug design 121, 186
ligand-protein pairwise 946, 950, 953
ligand sampling 804, 810, 813, 848-850, 853, 868
linear accelerator 691-692, 708
liposomes 340, 356, 384-385, 387-388, 391, 394
living donors 207, 210, 215
Lock-and-Key Theory 802
low intensity staffing model 727

M

machine learning 36, 38, 47-48, 64, 134, 420, 506, 607, 780, 783, 834, 861
malignant melanoma 542-543, 545, 550, 558-560
M-Band Wavelet Transform (M-Band WT) 519, 522, 524, 528-529, 532-533, 536
mechanism-based inhibitors 120, 132, 186

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

medical image fusion 519-524, 528, 532-533, 536
medical imaging 325, 327, 520-521, 545, 554
melanoma 303, 307-308, 366, 370-371, 381, 394, 542-546, 549-550, 554-555, 558-560, 563, 663
metabolic syndrome 218, 222, 233
metallic nanoparticles 312, 315-321, 384
methodological advances 807
mixed model 18
molecular descriptors 20-22, 26, 28, 31-32, 34-37, 40-41, 43, 48, 64, 86, 91-92, 149, 151, 413, 415, 420, 422, 424, 427-428, 437, 441-443, 457, 484, 490, 518, 575, 620, 624, 883, 903, 947
molecular docking 21, 95, 142, 437, 448, 451, 459-460, 491-492, 499, 501, 576-577, 593, 615, 618, 620, 623, 764-766, 771-776, 779, 781-783, 803-806, 809-810, 815, 817, 819, 829, 831-837, 840-842, 848-850, 856, 868-869, 876-880, 882-886, 891-893, 895, 902-904, 915-918, 924, 929, 941-942, 946, 948-950, 953-954
molecular dynamics 43, 93-94, 138, 142, 162, 494, 496-498, 510, 582, 606, 615-616, 765, 775, 778, 805, 809-810, 812, 818, 829, 831-832, 836-838, 841-842, 848, 856, 878-880, 882, 886, 897, 899, 903, 905, 926, 947
Molecular Dynamics (MD) simulation 138, 142, 497, 837-838, 841, 878-880, 882, 886
molecular properties 26, 64, 186, 412, 422, 446, 490, 607
molecular studies 371
mortality 1, 3-4, 7, 119, 121-122, 163, 204, 207, 216-218, 221-222, 229, 232, 238, 241, 246, 382, 506, 555, 559, 637-638, 650, 660-662, 667, 669, 690, 712-720, 739-740, 743, 750, 973
MTiAutoDock 941, 948, 953-954
multidisciplinary critical care 709

N

nanogenomics 367, 381
nanomaterials 95, 98, 311, 323, 325-326, 340, 343-344, 346-349, 351, 353, 355-356, 360, 366, 386, 394-396, 576, 902
nano materials 342-343
nanomedicine 321, 324, 340, 344, 346, 351, 356, 366, 369, 372, 374, 382, 384
nano medicine 342
nanometer (nm) 324, 353, 365
nanoparticles 95, 311-321, 323-326, 340, 342-344, 346, 349, 354-356, 360, 365, 369, 371, 374, 382, 384, 387-394, 407
nanoproteomics 367, 381
nanorobotics 324, 326
nanorods 356, 386
nanoscale 313, 324, 326-327, 340, 342-344, 346-347, 353-354, 361, 365, 367, 382-383, 495
nanoscience 311, 347, 352-353, 355, 365
nanosensor 340
nanoshells 344, 356, 384, 393
nanotechnology 311-314, 321, 324, 326-327, 340, 342-367, 381-384
nanotechnology particles 353, 355, 359-360, 365
neurological disorders 581
N-Glycolylneuraminic Acid (Neu5Gc) 236, 266
N-Nitroso Compounds 232-235, 265-266
Non-Communicable Diseases (NCD) 204, 638, 660-661, 666
non-maleficence 728-729, 738
non-steroidal aromatase inhibitors 118, 134, 145, 162
NS3 482, 484, 498-499, 501, 503, 507-508, 510-511
NS5B Polymerase 482, 490-493, 496-497, 506, 518, 782
Nurse Practitioners (NP) 709, 711-712

O

obesity 122, 216, 218, 228, 238, 240-242, 666, 704
omics 48, 374, 381
open ICU 727
organ transplants 203-204, 207, 210, 215

P

palliative care 1-10, 18-19, 187-192, 194, 197, 199, 205, 214, 647-648, 661, 665, 715, 720, 727, 733, 735-736, 749
palliative care consultation 2-4, 7-8, 18-19, 715
triggers 7-8, 19
PatchDock 862, 941, 948-952
PC-3 407
PEG 387, 389-390, 394, 407, 903
personalized medicine 385, 396, 579
pharmacology 21, 82, 85, 433, 574
pharmacophore 21, 26, 37, 40, 64, 94, 118, 133-135, 138, 143, 148, 152-154, 156-160, 162-163, 186, 426-428, 437, 447, 452, 484-487, 489, 493-494, 503-504, 511, 518, 593, 615, 618, 623, 817, 831, 833, 853, 892, 895, 898, 941
pharmacophore mapping 94, 118, 133-135, 143, 148, 156, 158-160, 162-163, 186, 426, 428, 593, 623
pharmacophore modeling 21, 138, 437, 485-487, 941
PHC centers 665-668
phenolic derivatives 409, 418, 425-426, 428
physical therapy 714-715

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Physician Assistants (PA) 709, 711-712
physician-patient communication 638, 742, 746, 751
PLA 388-389, 394, 407
plant-based 216-218, 238, 240, 245-246
point correspondence 697
Polycyclic Aromatic Hydrocarbons (PAH) 229, 233, 266
population health management 960-961, 963, 974
pose 9, 135, 139, 141, 206, 210, 388, 423, 449, 451-452, 463, 487, 493, 642, 772, 776, 781, 803, 810, 812, 814, 817, 832, 848, 853-856, 860-861, 868-869, 896, 906, 915-916, 922, 927-928, 952-954, 969
post-processing 31, 448, 451, 841-842, 848-849, 867, 869, 922, 926
pre-processing 32, 437, 542-543, 545-546, 561, 969
primary care 647, 662, 664, 666-669, 961
protein flexibility 486-487, 778, 782, 804, 808-810, 837, 848-849, 851-852, 856, 858, 868, 886, 899, 901-902, 917, 924
protein homology modeling 480
protein–ligand complexes 941
protein-protein docking 448, 571, 577-578, 590, 772, 776, 805, 848-849, 862-863, 868, 898-899
protein sequence alignment 460
PTX 395, 407

Q

quality assurance 703, 705, 963, 967
Quantitative Structure-Activity Relationship (QSAR) 20-21, 186, 409, 574, 590, 615, 813, 892, 941
quantum dots 325, 340, 354-355

R

radiation therapy (radiotherapy) 123, 372, 519-521, 664, 689-696, 700, 702-705, 708
receptor-ligand complex 133, 141
regenerative medicine 342, 346-347, 349, 351
relaxed complex scheme 829, 832, 842, 858
remediation 314, 340
right to health 637-638, 640-641, 739, 743, 750
risk communication 637-638, 640-641, 643, 649-650

S

sampling 43, 275, 526, 783, 804-805, 809-813, 832, 841-842, 848-856, 862-863, 868-869, 876, 878-879, 891, 893, 895, 898-901, 905, 916, 924, 928, 939
saturated fat 229-230, 242-245, 266
scoring 135, 141, 451, 486-487, 492-493, 501-502, 510-511, 562, 576, 578, 610, 764-765, 774, 776, 779-783, 803-805, 808-809, 812-814, 817, 832, 834-835, 840, 848-851, 853, 855-856, 859-863, 865, 868-869, 876-878, 880-883, 886, 891, 893, 895, 897-901, 906, 915-929, 939-940, 942, 947, 949, 953
scoring function 135, 451, 487, 501-502, 578, 779, 781, 804, 808, 812-814, 832, 834, 851, 855, 859-861, 865, 868-869, 878, 886, 895, 897-901, 906, 915-917, 919-924, 926-929, 942, 949, 953
scoring functions 135, 141, 451, 487, 492-493, 764-765, 776, 779-783, 803-804, 812-814, 817, 832, 834-835, 848-851, 853, 856, 859-862, 869, 876-878, 880-883, 886, 891, 895, 897, 906, 915-929, 939
searching function 808
search space 38, 449, 451, 775-776, 778, 782-783, 803, 853, 855, 897
segmentation 302-306, 308, 542-543, 545, 548-549, 559-562, 564, 566
self-created avatar 684, 686
SEM 393, 407
sensitivity 75, 121, 125, 230-231, 233, 240, 271, 273, 275-280, 283, 285, 287-288, 325, 349, 450, 486, 526, 545, 554-555, 560, 562-564, 640, 710-711, 747-748
sequence analysis 503, 571-573, 590
skin cancer 268, 302-303, 305, 542-546, 550-551, 555, 557-560, 562-564, 566, 663, 681, 683-684
social networking 676, 685-686
solvation 44, 94, 487, 497, 507, 615, 780, 812, 841, 850-851, 859, 862, 865, 869, 897, 916-918, 920-921, 924-929, 940, 952
solvation effect 897, 917, 924-926, 928
specialized palliative care 2, 4, 10, 19
specificity 121, 127, 133, 135, 162, 267, 271, 277, 279, 283, 285, 287-288, 353, 382-383, 387, 450, 486, 498, 554-555, 560, 563-564, 577, 594, 776, 917, 924-925, 927-929, 940
staffing models 709, 715, 717-718, 720
statistical methods 22, 64, 416, 420, 428, 433, 503, 549, 551, 838
stem cells 346, 351
storn 898
structure-based drug design 21, 46, 118, 121, 133, 186, 507, 511, 571, 577-578, 764-765, 772, 782-783, 803, 848-849, 851, 891, 893, 915-917, 942, 945
supercomputers 571, 580-582, 590, 885
superparamagnetic nanoparticles 355, 365
surgical buy-in 19

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

surgical critical illness 3, 5, 19
surrogate 94, 728, 731-733, 735-736, 738
SwissDock 577, 941, 948, 950, 952-953

T

team science 727
teamwork 711-712
tele-ICU 709, 718-720, 727
test data set 39, 64, 621, 623
Tf 231, 304, 384, 407
thanatourism 204, 215
tissue engineering 312, 326, 342, 347-349, 351, 354, 358, 363
training data set 38, 64, 93, 420, 612-613, 621, 623
trans-fats 231, 266
triaging 267, 269-270, 273, 278, 285, 288-289
Trimethylamine-N-Oxide (TMAO) 217, 235, 266
Tuberculosis 434, 436, 447, 453-454, 456, 459-463, 465, 580, 814, 902, 905
tumor 119, 122-127, 133, 268, 324, 344, 346, 370-371, 383-396, 521, 523, 536, 558, 581, 662, 690, 692-693, 708, 729, 864-865

U

Uveal Melanoma 366, 370, 381

V

vegan 216-217, 236, 238-240, 244-245
vegetarian 216-217, 235-236, 238-240, 244-245
virtual combinatorial library design 436, 443-444, 446, 480
virtual screening 46, 94, 118, 135, 142, 152-153, 162-163, 434, 436, 438-439, 447-448, 450, 480, 482-486, 489, 492-493, 496, 504, 509-512, 571, 573-574, 590, 593, 623, 772, 781-782, 813-814, 817, 819, 829-831, 833, 836, 842, 849, 856, 858, 860, 867, 876-878, 880, 884-886, 891-892, 902, 916, 924, 928-929, 940-942
virtual simulation 689, 691-692, 694, 708

W

women’s breast cancer awareness 739, 751

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use

Receive Free Lifetime E-Access or Free Hardcover*

Purchase a print book or e-book through the IGI Global Online Bookstore and receive the alternate version for free! Shipping fees apply.

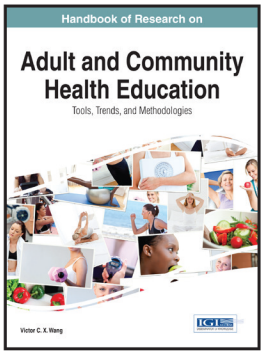
www.igi-global.com

Recommended Reference Books

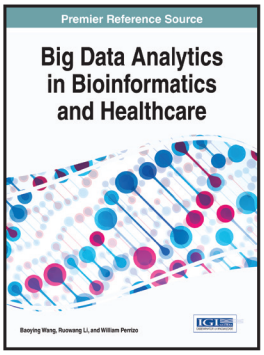
Take 20% Off through the IGI Global Online Bookstore**



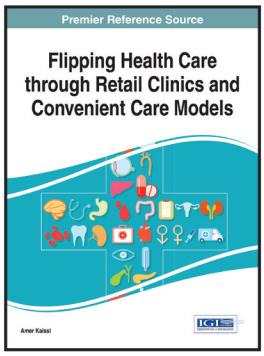
ISBN: 978-1-4666-4546-2
© 2014; 486 pp.
Take 20% Off:* $292



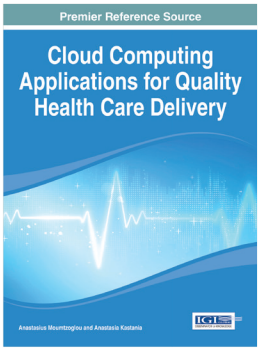
ISBN: 978-1-4666-6260-5
© 2014; 485 pp.
Take 20% Off:* $260



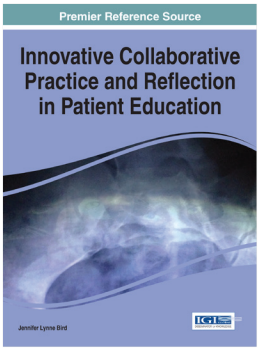
ISBN: 978-1-4666-6611-5
© 2015; 528 pp.
Take 20% Off:* $204



ISBN: 978-1-4666-6355-8
© 2015; 306 pp.
Take 20% Off:* $196



ISBN: 978-1-4666-6118-9
© 2014; 342 pp.
Take 20% Off:* $196



ISBN: 978-1-4666-7524-7
© 2015; 308 pp.
Take 20% Off:* $180

*IGI Global now offers the option to purchase hardcover, e-access, or hardcover + e-access for one price! You choose the format that best suits your needs. This offer is only valid on purchases made directly through IGI Global's Online Bookstore and not intended for use by book distributors or wholesalers. Shipping fees will be applied for hardcover purchases during checkout if this option is selected. E-Access will remain available for the lifetime of the publication and will only allow access to the edition purchased. Free Lifetime E-Access is only available to individuals and single institutions that purchase printed publications directly through IGI Global. Sharing the Free Lifetime E-Access is prohibited and will result in the termination of e-access. **20% discount cannot be combined with any other offer. Only valid on purchases made directly through IGI Global's Online Bookstore and not intended for use by book distributors or wholesalers. Not applicable on databases.

Publishing Progressive Information Science and Technology Research Since 1988

IGI GLOBAL
DISSEMINATOR OF KNOWLEDGE

www.igi-global.com | Sign up at www.igi-global.com/newsletters | facebook.com/igiglobal | twitter.com/igiglobal

EBSCOhost - printed on 2/11/2023 1:21 AM via . All use subject to https://www.ebsco.com/terms-of-use