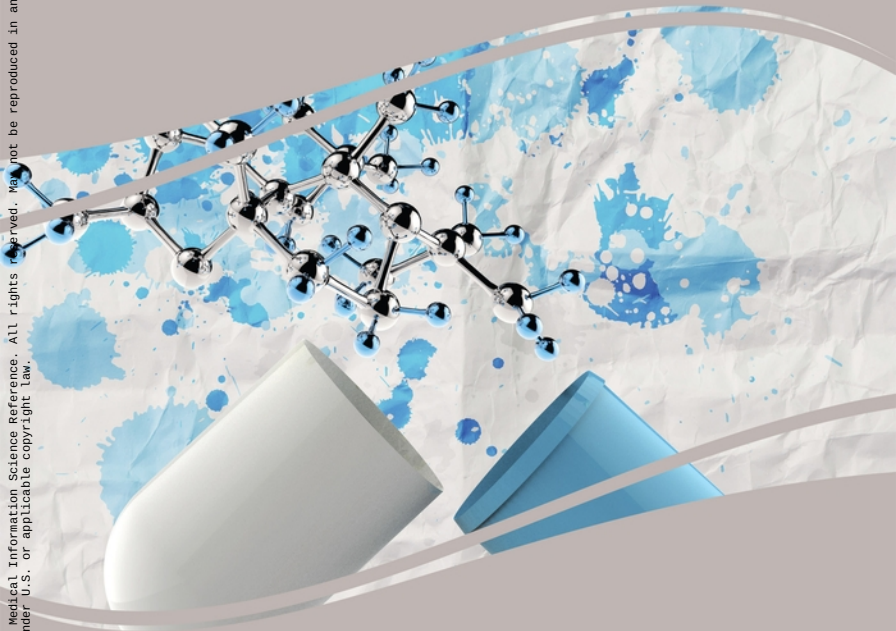
Premier Reference Source

Quantitative Structure-Activity Relationships in Drug Design, Predictive Toxicology, and Risk Assessment



Kunal Roy

Copyright 2015. 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 2/14/2023 7:16 AM via
AN: 961615 ; Roy, Kunal.; Quantitative Structure-Activity Relationships in Drug Design, Predictive Toxicology, and Risk Assessment
Account: ns335141

Quantitative Structure-Activity Relationships in Drug Design, Predictive Toxicology, and Risk Assessment

Kunal Roy
Jadavpur University, India

A volume in the Advances in Chemical and Materials Engineering (ACME) Book Series

Medical Information Science Reference
An Imprint of IGI Global

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Managing Director: Lindsay Johnston
Managing Editor: Austin DeMarco
Director of Intellectual Property & Contracts: Jan Travers
Acquisitions Editor: Kayla Wolfe
Production Editor: Christina Henning
Development Editor: Allison McGinniss
Typesetter: Amanda Smith
Cover Design: Jason Mull

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 © 2015 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

Quantitative structure-activity relationships in drug design, predictive toxicology, and risk assessment / Kunal Roy, editor.
p. ; cm.
Includes bibliographical references and index.
Summary: “This book discusses recent advancements in the field of QSARs with special reference to their application in drug development, predictive toxicology, and chemical risk analysis”--Provided by publisher.
ISBN 978-1-4666-8136-1 (hardcover) -- ISBN 978-1-4666-8137-8 (ebook)
I. Roy, Kunal, 1971- , editor.
[DNLM: 1. Drug Design. 2. Quantitative Structure-Activity Relationship. 3. Drug Antagonism. 4. Drug-Related Side Effects and Adverse Reactions. 5. Risk Assessment. QV 745]
RS420
615.1’9--dc23
2015000076

This book is published in the IGI Global book series Advances in Chemical and Materials Engineering (ACME) (ISSN: 2327-5448; eISSN: 2327-5456)

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 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use



Advances in Chemical and Materials Engineering (ACME) Book Series

J. Paulo Davim
University of Aveiro, Portugal

ISSN: 2327-5448
EISSN: 2327-5456

Mission

The cross disciplinary approach of chemical and materials engineering is rapidly growing as it applies to the study of educational, scientific and industrial research activities by solving complex chemical problems using computational techniques and statistical methods.

The **Advances in Chemical and Materials Engineering (ACME) Book Series** provides research on the recent advances throughout computational and statistical methods of analysis and modeling. This series brings together collaboration between chemists, engineers, statisticians, and computer scientists and offers a wealth of knowledge and useful tools to academics, practitioners, and professionals through high quality publications.

Coverage

- Wear of Materials
- Industrial Chemistry
- Fracture Mechanics
- Polymers Engineering
- Biomaterials
- Computational Methods
- Electrochemical and Corrosion
- Toxicity of Materials
- Ceramics and Glasses
- Materials to Renewable Energies

IGI Global is currently accepting manuscripts for publication within this series. To submit a proposal for a volume in this series, please contact our Acquisition Editors at Acquisitions@igi-global.com or visit: http://www.igi-global.com/publish/.

The Advances in Chemical and Materials Engineering (ACME) Book Series (ISSN 2327-5448) is published by IGI Global, 701 E. Chocolate Avenue, Hershey, PA 17033-1240, USA, www.igi-global.com. This series is composed of titles available for purchase individually; each title is edited to be contextually exclusive from any other title within the series. For pricing and ordering information please visit http://www.igi-global.com/book-series/advances-chemical-materials-engineering/73687. Postmaster: Send all address changes to above address. Copyright © 2015 IGI Global. All rights, including translation in other languages reserved by the publisher. No part of this series may be reproduced or used in any form or by any means – graphics, electronic, or mechanical, including photocopying, recording, taping, or information and retrieval systems – without written permission from the publisher, except for non commercial, educational use, including classroom teaching purposes. The views expressed in this series are those of the authors, but not necessarily of IGI Global.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Titles in this Series

For a list of additional titles in this series, please visit: www.igi-global.com

Processing Techniques and Tribological Behavior of Composite Materials
Rajnesh Tyagi (Indian Institute of Technology (BHU), Varanasi, India) and J. Paulo Davim (University of Aveiro, Portugal)
Engineering Science Reference • copyright 2015 • 410pp • H/C (ISBN: 9781466675308) • US $225.00 (our price)

Handbook of Research on Diverse Applications of Nanotechnology in Biomedicine, Chemistry, and Engineering
Shivani Soni (Alabama State University, USA) Amandeep Salhotra (City of Hope National Medical Center, USA) and Mrutyunjay Suar (KIIT University, India)
Engineering Science Reference • copyright 2015 • 820pp • H/C (ISBN: 9781466663633) • US $395.00 (our price)

Innovative Applications and Developments of Micro-Pattern Gaseous Detectors
Tom Francke (Myon, Sweden) and Vladimir Peskov (CERN, Switzerland)
Engineering Science Reference • copyright 2014 • 303pp • H/C (ISBN: 9781466660144) • US $215.00 (our price)

Handbook of Research on Nanoscience, Nanotechnology, and Advanced Materials
Mohamed Bououdina (University of Bahrain, Bahrain) and J. Paulo Davim (University of Aveiro, Portugal)
Engineering Science Reference • copyright 2014 • 617pp • H/C (ISBN: 9781466658240) • US $365.00 (our price)

Quantum and Optical Dynamics of Matter for Nanotechnology
Mihai V. Putz (West University of Timisoara, Romania)
Engineering Science Reference • copyright 2014 • 527pp • H/C (ISBN: 9781466646872) • US $180.00 (our price)

Advanced Solar Cell Materials, Technology, Modeling, and Simulation
Laurentiu Fara (Polytechnic University of Bucharest, Romania) and Masafumi Yamaguchi (Toyota Technological Institute, Japan)
Engineering Science Reference • copyright 2013 • 354pp • H/C (ISBN: 9781466619272) • US $195.00 (our price)

Computational Gas-Solids Flows and Reacting Systems Theory, Methods and Practice
Sreekanth Pannala (Oak Ridge National Laboratory, USA) Madhava Syamlal (National Energy Technology Laboratory, USA) and Thomas J. O'Brien (National Energy Technology Laboratory, USA)
Engineering Science Reference • copyright 2011 • 500pp • H/C (ISBN: 9781615206513) • US $180.00 (our price)

IGI GLOBAL
DISSEMINATOR OF KNOWLEDGE
www.igi-global.com

701 E. Chocolate Ave., Hershey, PA 17033
Order online at www.igi-global.com or call 717-533-8845 x100
To place a standing order for titles released in this series, contact: cust@igi-global.com
Mon-Fri 8:00 am - 5:00 pm (est) or fax 24 hours a day 717-533-8661

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Editorial Advisory Board

Emilio Benfenati, *Istituto di Ricerche Farmacologiche Mario Negri, Italy*
Eduardo A. Castro, *La Plata University, Argentina*
Asim K. Debnath, *New York Blood Center, USA*
S. P. Gupta, *National Institute of Technical Teachers' Training and Research (NITTTR), India*
Lemont B. Kier, *Virginia Commonwealth University, USA*
Jerzy Leszczynski, *Jackson State University, USA*
Bono Lučić, *Ruđer Bošković Institute, Croatia*
Paul L. A. Popelier, *Manchester Institute of Biotechnology, UK*

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table of Contents

Foreword xvi

Preface xviii

Acknowledgment xxiv

Chapter 1
An Introduction to the Basic Concepts in QSAR-Aided Drug Design 1
Maryam Hamzeh-Mivehroud, Biotechnology Research Center & School of Pharmacy, Tabriz University of Medical Sciences, Tabriz, Iran
Babak Sokouti, Biotechnology Research Center, Tabriz University of Medical Sciences, Tabriz, Iran
Siavoush Dastmalchi, Biotechnology Research Center & School of Pharmacy, Tabriz University of Medical Sciences, Tabriz, Iran

Chapter 2
The "ETA" Indices in QSAR/QSPR/QSTR Research 48
Kunal Roy, Jadavpur University, India
Rudra Narayan Das, Jadavpur University, India

Chapter 3
Evolution of Multivariate Image Analysis in QSAR: The Case for a Neglected Disease 84
Matheus P. Freitas, Federal University of Lavras, Brazil
Mariene H. Duarte, Federal University of Lavras, Brazil

Chapter 4
Quantitative Structure-Activity/Property/Toxicity Relationships through Conceptual Density Functional Theory-Based Reactivity Descriptors 123
Sudip Pan, Indian Institute of Technology Kharagpur, India
Ashutosh Gupta, Udai Pratap Autonomous College, India
Venkatesan Subramanian, Central Leather Research Institute, India
Pratim K. Chattaraj, Indian Institute of Technology Kharagpur, India

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 5
Importance of Applicability Domain of QSAR Models 180
Kunal Roy, Jadavpur University, India
Supratik Kar, Jadavpur University, India

Chapter 6
QSAR of Antioxidants 212
Omar Deeb, Al-Quds University, Palestine
Mohammad Goodarzi, Katholieke Universiteit Leuven, Belgium

Chapter 7
QSAR Studies on Bacterial Efflux Pump Inhibitors 238
Khac-Minh Thai, University of Medicine and Pharmacy at HCMC, Vietnam
Trong-Nhat Do, University of Medicine and Pharmacy at HCMC, Vietnam
Thuy-Viet-Phuong Nguyen, University of Medicine and Pharmacy at HCMC, Vietnam
Duc-Khanh-Tho. Nguyen, University of Medicine and Pharmacy at HCMC, Vietnam
Thanh-Dao Tran, University of Medicine and Pharmacy at HCMC, Vietnam

Chapter 8
Integrated *in Silico* Methods for the Design and Optimization of Novel Drug Candidates: A Case Study on Fluoroquinolones – *Mycobacterium tuberculosis* DNA Gyrase Inhibitors 269
Nikola Minovski, National Institute of Chemistry, Slovenia
Marjana Novič, National Institute of Chemistry, Slovenia

Chapter 9
Computational Approaches for the Discovery of Novel Hepatitis C Virus NS3/4A and NS5B Inhibitors 318
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 10
QSAR Models towards Cholinesterase Inhibitors for the Treatment of Alzheimer's Disease 354
C. Gopi Mohan, Amrita Institute of Medical Sciences and Research Centre, India
Shikhar Gupta, National Institute of Pharmaceutical Education and Research, India

Chapter 11
Ligand- and Structure-Based Drug Design of Non-Steroidal Aromatase Inhibitors (NSAIs) in Breast Cancer 400
Tarun Jha, Jadavpur University, India
Nilanajn Adhikari, Jadavpur University, India
Amit Kumar Halder, Jadavpur University, India
Achintya Saha, University of Calcutta, India

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 12
Computational Techniques Application in Environmental Exposure Assessment 471
Karolina Jagiello, University of Gdansk, Poland
Tomasz Puzyn, University of Gdansk, Poland

Chapter 13
QSAR-Based Studies of Nanomaterials in the Environment 506
Valeria V. Kleandrova, University of Porto, Portugal
Feng Luan, Yantai University, China & University of Porto, Portugal
Alejandro Speck-Planche, University of Porto, Portugal
M. Natália D. S. Cordeiro, University of Porto, Portugal

Chapter 14
Quantitative Nanostructure-Activity Relationship Models for the Risk Assessment of NanoMaterials 535
Eleni Vrontaki, NovaMechanics Ltd., Cyprus & University of Athens, Greece
Thomas Mavromoustakos, University of Athens, Greece
Georgia Melagraki, NovaMechanics Ltd., Cyprus
Antreas Afantitis, NovaMechanics Ltd., Cyprus

Chapter 15
QSPR/QSAR Analyses by Means of the CORAL Software: Results, Challenges, Perspectives 560
Andrey A. Toropov, IRCCS-Istituto di Ricerche Farmacologiche Mario Negri, Italy
Alla P. Toropova, IRCCS-Istituto di Ricerche Farmacologiche Mario Negri, Italy
Emilio Benfenati, IRCCS-Istituto di Ricerche Farmacologiche Mario Negri, Italy
Orazio Nicolotti, Università degli Studi di Bari "Aldo Moro", Italy
Angelo Carotti, Università degli Studi di Bari "Aldo Moro", Italy
Karel Nesmerak, Charles University in Prague, Czech Republic
Aleksandar M. Veselinović, University of Niš, Serbia
Jovana B. Veselinović, University of Niš, Serbia
Pablo R. Duchowicz, Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas INIFTA (UNLP, CCT La Plata-CONICET), Argentina
Daniel Bacelo, Universidad de Belgrano, Argentina
Eduardo A. Castro, Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas INIFTA (UNLP, CCT La Plata-CONICET), Argentina
Bakhtiyor F. Rasulev, Jackson State University, USA
Danuta Leszczynska, Jackson State University, USA
Jerzy Leszczynski, Jackson State University, USA

Compilation of References 586

About the Contributors 688

Index 701

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Detailed Table of Contents

Foreword xvi

Preface xviii

Acknowledgment xxiv

Chapter 1
An Introduction to the Basic Concepts in QSAR-Aided Drug Design 1
Maryam Hamzeh-Mivehroud, Biotechnology Research Center & School of Pharmacy, Tabriz University of Medical Sciences, Tabriz, Iran
Babak Sokouti, Biotechnology Research Center, Tabriz University of Medical Sciences, Tabriz, Iran
Siavoush Dastmalchi, Biotechnology Research Center & School of Pharmacy, Tabriz University of Medical Sciences, Tabriz, Iran

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.

Chapter 2
The "ETA" Indices in QSAR/QSPR/QSTR Research 48
Kunal Roy, Jadavpur University, India
Rudra Narayan Das, Jadavpur University, India

Descriptors are one of the most essential components of predictive Quantitative Structure-Activity/Property/Toxicity Relationship (QSAR/QSPR/QSTR) modeling analysis, as they encode chemical information of molecules in the form of quantitative numbers, which are used to develop mathematical correlation models. The quality of a predictive model not only depends on good modeling statistics, but also on the extraction of chemical features. A significant amount of research since the beginning of QSAR analysis paradigm has led to the introduction of a large number of predictor variables or descriptors. The

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Extended Topochemical Atom (ETA) indices, developed by the authors' group, successfully address the aspects of molecular topology, electronic information, and different types of bonded interactions, and have been extensively employed for the modeling of different types of activity/property and toxicity endpoints. This chapter provides explicit information regarding the basis, algorithm, and applicability of the ETA indices for a predictive modeling paradigm.

Chapter 3
Evolution of Multivariate Image Analysis in QSAR: The Case for a Neglected Disease 84
Matheus P. Freitas, Federal University of Lavras, Brazil
Mariene H. Duarte, Federal University of Lavras, Brazil

Multivariate Image Analysis applied in Quantitative Structure-Activity Relationship (MIA-QSAR) is a simple method to achieve, at least in a variety of examples, QSAR models with predictive abilities comparable to those of sophisticated tridimensional methodologies. MIA-QSAR is based on the correlation between properties (e.g. biological activities) and chemical descriptors, which are pixels of images representing chemical structures in a congeneric series of molecules. The MIA-QSAR approach has been improved since its creation, in 2005, both in terms of data analysis and development of more descriptive information. This chapter reports the MIA-QSAR method, including its augmented version, named aug-MIA-QSAR because of the introduction of new dimensions to better encode atomic properties. In addition, the application to a case study illustrates the main practical differences between traditional and augmented MIA-QSAR. The use of a neglected disease as example represents a challenge in QSAR, which is particularly focused on diseases with higher economical appearance.

Chapter 4
Quantitative Structure-Activity/Property/Toxicity Relationships through Conceptual Density Functional Theory-Based Reactivity Descriptors 123
Sudip Pan, Indian Institute of Technology Kharagpur, India
Ashutosh Gupta, Udai Pratap Autonomous College, India
Venkatesan Subramanian, Central Leather Research Institute, India
Pratim K. Chattaraj, Indian Institute of Technology Kharagpur, India

Developing effective structure-activity/property/toxicity relationships (QSAR/QSPR/QSTR) is very helpful in predicting biological activity, property, and toxicity of a given set of molecules. Regular change in these properties with the structural alteration is the main reason to obtain QSAR/QSPR/QSTR models. The advancement in making different QSAR/QSPR/QSTR models to describe activity, property, and toxicity of various groups of molecules is reviewed in this chapter. The successful implementation of Conceptual Density Functional Theory (CDFT)-based global as well as local reactivity descriptors in modeling effective QSAR/QSPR/QSTR is highlighted.

Chapter 5
Importance of Applicability Domain of QSAR Models 180
Kunal Roy, Jadavpur University, India
Supratik Kar, Jadavpur University, India

Quantitative Structure-Activity Relationship (QSAR) models have manifold applications in drug discovery, environmental fate modeling, risk assessment, and property prediction of chemicals and pharmaceuticals. One of the principles recommended by the Organization of Economic Co-operation and Development

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

(OECD) for model validation requires defining the Applicability Domain (AD) for QSAR models, which allows one to estimate the uncertainty in the prediction of a compound based on how similar it is to the training compounds, which are used in the model development. The AD is a significant tool to build a reliable QSAR model, which is generally limited in use to query chemicals structurally similar to the training compounds. Thus, characterization of interpolation space is significant in defining the AD. An attempt is made in this chapter to address the important concepts and methodology of the AD as well as criteria for estimating AD through training set interpolation in the descriptor space.

Chapter 6
QSAR of Antioxidants 212
Omar Deeb, Al-Quds University, Palestine
Mohammad Goodarzi, Katholieke Universiteit Leuven, Belgium

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.

Chapter 7
QSAR Studies on Bacterial Efflux Pump Inhibitors 238
Khac-Minh Thai, University of Medicine and Pharmacy at HCMC, Vietnam
Trong-Nhat Do, University of Medicine and Pharmacy at HCMC, Vietnam
Thuy-Viet-Phuong Nguyen, University of Medicine and Pharmacy at HCMC, Vietnam
Duc-Khanh-Tho. Nguyen, University of Medicine and Pharmacy at HCMC, Vietnam
Thanh-Dao Tran, University of Medicine and Pharmacy at HCMC, Vietnam

Antimicrobial drug resistance occurs when bacteria undergo certain modifications to eliminate the effectiveness of drugs, chemicals, or other agents designed to cure infections. To date, the burden of resistance has remained one of the major clinical concerns as it renders prolonged and complicated treatments, thereby increasing the medical costs with lengthier hospital stays. Of complex causes for bacterial resistance, there has been increasing evidence that proved the significant role of efflux pumps in antibiotic resistance. Coadministration of Efflux Pump Inhibitors (EPIs) with antibiotics has been considered one of the promising ways not only to improve the efficacy but also to extend the clinical utility of existing antibiotics. This chapter begins with outlining current knowledge about bacterial efflux pumps and drug designs applied in identification of their modulating compounds. Following, the chapter addresses and provides a discussion on Quantitative Structure-Activity Relationship (QSAR) analyses in search of novel and potent efflux pump inhibitors.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 8
Integrated *in Silico* Methods for the Design and Optimization of Novel Drug Candidates: A Case Study on Fluoroquinolones – *Mycobacterium tuberculosis* DNA Gyrase Inhibitors 269
Nikola Minovski, National Institute of Chemistry, Slovenia
Marjana Novič, National Institute of Chemistry, Slovenia

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.

Chapter 9
Computational Approaches for the Discovery of Novel Hepatitis C Virus NS3/4A and NS5B Inhibitors 318
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

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.

Chapter 10
QSAR Models towards Cholinesterase Inhibitors for the Treatment of Alzheimer's Disease 354
C. Gopi Mohan, Amrita Institute of Medical Sciences and Research Centre, India
Shikhar Gupta, National Institute of Pharmaceutical Education and Research, India

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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.

Chapter 11
Ligand- and Structure-Based Drug Design of Non-Steroidal Aromatase Inhibitors (NSAIs) in Breast Cancer 400
Tarun Jha, Jadavpur University, India
Nilanajn Adhikari, Jadavpur University, India
Amit Kumar Halder, Jadavpur University, India
Achintya Saha, University of Calcutta, India

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.

Chapter 12
Computational Techniques Application in Environmental Exposure Assessment 471
Karolina Jagiello, University of Gdansk, Poland
Tomasz Puzyn, University of Gdansk, Poland

In this chapter, the application of computational techniques in environmental exposure assessment was described. The most important groups of these techniques are Multimedia Mass-balance (MM) modelling and Quantitative Structure-Activity/Structure-Property Relationships (QSAR/QSPR) modelling. Multimedia Mass-balance models have been widely utilized for studying Long-Range Transport Potential (LRTP) and overall persistence (POV) of Persistent Organic Pollutants (POPs), regulated by many national and international acts, including the Stockholm Convention on POPs. Recently, a novel modelling methodology that links QSPR and MM has been implemented. According to this approach, the physical/chemical properties required as the input variables for multimedia modelling can be calculated directly from appropriate QSPR models. QSPR models must be previously developed based on the relationships between the chemical structure and the modelled properties (QSPR).

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 13
QSAR-Based Studies of Nanomaterials in the Environment 506
Valeria V. Kleandrova, University of Porto, Portugal
Feng Luan, Yantai University, China & University of Porto, Portugal
Alejandro Speck-Planche, University of Porto, Portugal
M. Natália D. S. Cordeiro, University of Porto, Portugal

Nanotechnology is a newly emerging field, posing substantial impacts on society, economy, and the environment. In recent years, the development of nanotechnology has led to the design and large-scale production of many new materials and devices with a vast range of applications. However, along with the benefits, the use of nanomaterials raises many questions and generates concerns due to the possible health-risks and environmental impacts. This chapter provides an overview of the Quantitative Structure-Activity Relationships (QSAR) studies performed so far towards predicting nanoparticles' environmental toxicity. Recent progresses on the application of these modeling studies are additionally pointed out. Special emphasis is given to the setup of a QSAR perturbation-based model for the assessment of ecotoxic effects of nanoparticles in diverse conditions. Finally, ongoing challenges that may lead to new and exciting directions for QSAR modeling are discussed.

Chapter 14
Quantitative Nanostructure-Activity Relationship Models for the Risk Assessment of NanoMaterials 535
Eleni Vrontaki, NovaMechanics Ltd., Cyprus & University of Athens, Greece
Thomas Mavromoustakos, University of Athens, Greece
Georgia Melagraki, NovaMechanics Ltd., Cyprus
Antreas Afantitis, NovaMechanics Ltd., Cyprus

In the last few decades, nanotechnology has been deeply established into human's everyday life with a great number of applications in cosmetics, textiles, electronics, optics, medicine, and many more. Although nanotechnology applications are rapidly increasing, the toxicity of some nanomaterials to living organisms and the environment still remains unknown and needs to be explored. The traditional toxicological evaluation of nanoparticles with the wide range of types, shapes, and sizes often involves expensive and time-consuming procedures. An efficient and cheap alternative is the development and application of predictive computational models using Quantitative Nanostructure-Activity Relationship (QNAR) methods. Towards this goal, researchers are mainly focused on the adverse effects of metal oxides and carbon nanotubes, but to date, QNAR studies are rare mainly because of the limited number of available organized datasets. In this chapter, recent studies for predictive QNAR models for the risk assessment of nanomaterials are reported and the perspectives of computational nanotoxicology that deeply relies on the intense collaboration between experimental and computational scientists are discussed.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 15
QSPR/QSAR Analyses by Means of the CORAL Software: Results, Challenges, Perspectives 560
Andrey A. Toropov, IRCCS-Istituto di Ricerche Farmacologiche Mario Negri, Italy
Alla P. Toropova, IRCCS-Istituto di Ricerche Farmacologiche Mario Negri, Italy
Emilio Benfenati, IRCCS-Istituto di Ricerche Farmacologiche Mario Negri, Italy
Orazio Nicolotti, Università degli Studi di Bari "Aldo Moro", Italy
Angelo Carotti, Università degli Studi di Bari "Aldo Moro", Italy
Karel Nesmerak, Charles University in Prague, Czech Republic
Aleksandar M. Veselinović, University of Niš, Serbia
Jovana B. Veselinović, University of Niš, Serbia
Pablo R. Duchowicz, Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas INIFTA (UNLP, CCT La Plata-CONICET), Argentina
Daniel Bacelo, Universidad de Belgrano, Argentina
Eduardo A. Castro, Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas INIFTA (UNLP, CCT La Plata-CONICET), Argentina
Bakhtiyor F. Rasulev, Jackson State University, USA
Danuta Leszczynska, Jackson State University, USA
Jerzy Leszczynski, Jackson State University, USA

In this chapter, the methodology of building up quantitative structure—property/activity relationships (QSPRs/QSARs)—by means of the CORAL software is described. The Monte Carlo method is the basis of this approach. Simplified Molecular Input-Line Entry System (SMILES) is used as the representation of the molecular structure. The conversion of SMILES into the molecular graph is available for QSPR/QSAR analysis using the CORAL software. The model for an endpoint is a mathematical function of the correlation weights for various features of the molecular structure. Hybrid models that are based on features extracted from both SMILES and a graph also can be built up by the CORAL software. The conceptually new ideas collected and revealed through the CORAL software are: (1) any QSPR/QSAR model is a random event; and (2) optimal descriptor can be a translator of eclectic information into an endpoint prediction.

Compilation of References 586

About the Contributors 688

Index 701

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Foreword

From the title of this book, we capture the broad extent of this scientific paradigm built around the drug molecule. This book successfully records the position of drug research, now in its third generation. To reminisce a moment, we go back to the first generation, the beginning of mathematics and structure description introduced into the current paradigm. The decades of the 1960s and 1970s witnessed the origin of mathematics applied to molecular structure with the intent to be predictive in drug design. We speak of QSAR. The origin of the paradigm in the first generation was built around physical properties first, followed by numerical descriptors of structure. Included was the use of molecular orbital theory to encode confirmation, the use of topology to introduce a vast area of structure description, and the identification of a critical structure feature on an active molecule, called the pharmacophore. These were the introductory elements in the 1960s and 1970s that underlie where we are today.

A second generation emerged in the 1980s and 1990s, where broad acceptance of these approaches came into being in academia and industry. Accompanying this acceptance was the development of a large number of methods seeking information necessary for understanding and drug design. Various modelling paradigms were introduced leading to insight guiding drug design. With the accompanying gain in mathematical theories and computer methods, this second generation led to highly successful strategies that greatly enriched successful drug design.

We are now in the third generation of this grand paradigm. It is characterized by a significant broadening of the approaches developed earlier. Look at the titles of the articles written by the 50 authors here. They reflect interest and successful development of research far beyond just drug design. There is predictive toxicology, risk assessment, nanomaterials, antioxidants, just to name a few. This third generation in 2000-2020 is one of immense dispersion of attention and applications of the paradigm developed earlier.

This book captures the essence of this third generation. It should play an active part in the education of everyone pursuing the goals that compose this generation of scientists involved in the subjects covered in this splendid text, expertly assembled by Kunal Roy.

Lemont B. Kier
Virginia Commonwealth University, USA

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Lemont B. Kier *is an Emeritus Professor of Medicinal Chemistry at the Virginia Commonwealth University, Richmond, Virginia, USA. Dr. Kier received a BS in Pharmacy from The Ohio State University, USA, and a PhD in Medicinal Chemistry from the University of Minnesota, USA. Dr. Kier was an early developer of drug design strategies using quantitative structure-activity information. This included molecular orbital theory to predict preferred conformations and numerical descriptors encoding molecular structure and electronic properties. More recent research has developed new approaches to the modeling of molecular structure using cellular automata. These include studies of dynamic processes leading to theories of anesthetic action, sleep, aging, and nerve conduction. Close to 300 papers and 8 books have revealed these contributions. Dr. Kier is active in medicinal chemistry, life sciences, and nurse anesthesia courses at the Virginia Commonwealth University.*

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Preface

Quantitative Structure-Activity Relationships (QSARs) represent predictive models derived from application of statistical tools correlating biological activity (including therapeutic and toxic) or other properties of chemicals (drugs/toxicants/environmental pollutants) with descriptors representative of molecular structure and/or property. QSAR is a sub-discipline of Cheminformatics, which has long been used in medicinal chemistry for lead optimization and drug design. It is basically a ligand-based statistical approach. However, with advancement in the knowledge of the target receptor structure information and related studies, the QSAR studies of the present days are usually performed in combination with the structure-based approaches leading to a good number of success stories. The knowledge of QSAR in association with other ligand-based approaches like pharmacophore mapping and structure-based approaches like docking coupled with virtual screening of compound libraries may identify novel hit compounds. QSAR increases the probability of success of finding an optimum lead with desired therapeutic efficacy and pharmacokinetic profile, increased selectivity, and minimum side effects at the same time, thus avoiding costly experiments with thousands of compounds with less potential to become successful, hence avoiding colossal expenditure. QSAR is very helpful in modeling absorption, distribution, metabolism, elimination, and toxicity profile of drug candidates in the initial phases of drug discovery. QSAR is also a very popular tool for risk assessment of chemicals in the absence of experimental data. Such approach is used by the United States Environmental Protection Agency and also encouraged in the European Union's REACH legislation. QSAR techniques are in consonance with the "3R concept" related to the moral principle regarding the use of sentient animals. The Organization of Economic Cooperation and Development (OECD) has recommended a set of guidelines for development and validation of QSAR models for regulatory uses. Apart from the conventional uses in drug discovery and predictive risk assessment, QSAR has also been used for modeling chemicals with special applications such as antioxidants, ionic liquids, and nanomaterials. In the present day QSAR research, emphasis is given on the (external) predictive quality and mechanistic interpretability of the models. However, it may be remembered that the best predictive model may not have sufficient mechanistic interpretability, and on the other hand, a model with very good interpretation may have low predictive ability. Thus, selection of the best model depends on the objective of the analysis besides other factors.

QSARs and related chemometric tools are extensively used in different fields of chemistry with applications in medicine, environment, and agriculture for ranking of potential compounds for prioritizing experiments. Starting in early 1960s with the Classical Hansch analysis and Free-Wilson model, QSAR has come a long way with much advancement and modernization with respect to the descriptors and validation tools used leading to evolution of a sophisticated scientific discipline on its own at the interface of chemistry, biology, and statistics.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

This book is intended to showcase the recent developments in the field of QSAR with its applications in different emerging fields. There are a total of 15 chapters on different topics contributed by the experts working in the field.

Before knowing the applications of QSAR in different fields of chemical research, the novice readers should have some knowledge on the preliminaries of QSAR tools. QSAR essentially being a statistical approach, a working knowledge of statistics is necessary for the development, validation, and interpretation of QSAR models. The first chapter entitled "An Introduction to the Basic Concepts in QSAR-Aided Drug Design" and contributed by Maryam Hamzeh-Mivehroud, Babak Sokouti, and Siavoush Dastmalchi covers the basic principles and tools used in QSAR. The authors give the background and different applications of QSAR followed by the introduction of molecular structure representation and the concept of descriptors. Then they discuss different descriptors and linear and non-linear statistical tools used in QSAR studies. Different validation strategies used in QSAR are also mentioned. The authors finally discuss the 3D-QSAR studies and higher dimensional QSARs. At the end, a few case studies are presented and future directions are indicated. This chapter should be a good starting point for those who are new to the field of QSAR.

Descriptors are fundamental to the development of QSAR models. They encode chemical information in a numerical form making it suitable for statistical operations. Though there are hundreds of descriptors present for development of QSAR models, there is always a need to develop novel descriptors that may address some of the deficiencies of the previous ones. Moreover, choice of a particular type of descriptor depends on the structural complexity present in the data set and the type of response being modeled. The second chapter entitled "The 'ETA' Indices in QSAR/QSPR/QSTR Research" contributed by Kunal Roy and Rudra Narayan Das presents the algorithm of the novel Extended Topochemical Atom (ETA) indices and their application in quantitative structure-activity/property/toxicity relationship studies. The authors describe different kinds of descriptors that are in use in QSAR studies and then indicate the advantages of two-dimensional descriptors. Then the evolution of the ETA descriptors is discussed, and the definitions of different variants of first and second generations of the ETA indices are elaborated. The authors also show sample calculation of the ETA indices for two representative molecules. The available software tools for the computation of the ETA indices are mentioned, and finally, applications of these indices in the modeling studies are discussed.

Multivariate Image Analysis (MIA) descriptors for QSAR are discussed in Chapter 3 ("Evolution of Multivariate Image Analysis in QSAR: The Case for a Neglected Disease") contributed by Matheus P. Freitas and Mariene H. Duarte. The authors explain the basic prociples of MIA-QSAR and augmented MIA-QSAR. They apply this approach for modeling a data set of compounds active against a neglected disease (trypanosomiasis). The future prospects of this approach are also discussed.

The fourth chapter entitled "Quantitative Structure-Activity/Property/Toxicity Relationships through Conceptual Density Functional Theory-Based Reactivity Descriptors" and contributed by Sudip Pan, Ashutosh Gupta, Venkatesan Subramanian, and Pratim K. Chattaraj discusses implementation of Conceptual Density Functional Theory (CDFT)-based descriptors in quantitative structure-activity/property/toxicity studies. The authors mention different types of descriptors that are used in QSAR analyses. Then they discuss some descriptors defined based on the CDFT. Examples are cited for the use of the CDFT-based descriptors in various modeling studies.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

The Applicability Domain (AD) of QSAR models is an important issue to check while predicting a new set of compounds using previously developed QSAR models. AD allows one to estimate the uncertainty in the prediction of a particular compound based on how similar it is to the training compounds that are used in the model development. Chapter 5 entitled "Importance of Applicability Domain of QSAR Models" and contributed by Kunal Roy and Supratik Kar defines the concept of applicability domain of QSAR models. Different approaches of determining AD of QSAR models (based on ranges in the descriptor space, geometry, distance, probability density distribution, and range of the response variable) are elaborated. Some practical examples of the determination of AD are also added. The future research direction is indicated.

The application of QSAR has been applied to different chemical classes of diverse applications. Antioxidants are an important class of chemicals with great bio-medicinal importance. Chapter 6 ("QSAR of Antioxidants"), contributed by Omar Deeb and Mohammad Goodarzi, discusses the applications of QSAR in modeling of antioxidant compounds. This chapter starts with an introductory discussion on antioxidants, their mechanism of action, and methods of determination of antioxidant activity. Then the concepts of QSAR, descriptors and statistical methods and validation strategies are briefly reviewed. Finally, the authors discuss the recent QSAR reports on antioxidants and comment on the structure-antioxidant activity relationships.

Though QSAR was originally developed in the context of physical organic chemistry, it has been applied very extensively to chemicals (drugs) with action on the biological systems. The applications of QSAR in medicinal chemistry have addressed design and development of potential ligands active against different diseases and health problems. The need for development of new drugs may arise due to emergence of new diseases or reduction of therapeutic efficacy of older candidates in a particular disease condition of increased complexity or emergence of drug resistance. Development of resistance of bacterial pathogens to various antimicrobial agents is an issue of high concern. One of the primary mechanisms of drug resistance is extrusion of the foreign chemical through bacterial efflux pump. To fight against multidrug efflux systems, one possible practice is the combination of conventional antimicrobial agents with small molecules known as multidrug efflux pump inhibitors. Chapter 7 ("QSAR Studies on Bacterial Efflux Pump Inhibitors"), contributed by Khac-Minh Thai, Trong-Nhat Do, Thuy-Viet-Phuong Nguyen, Khanh-Tho D Nguyen, and Thanh-Dao Tran, deals with QSAR modeling studies for bacterial efflux pump inhibitors. The chapter starts with an introduction on antibiotic resistance and role of bacterial efflux pumps in development of the resistance. Then various bacterial efflux pump inhibitors and targets for designing new efflux pump inhibitors are discussed. Finally, QSAR approaches for the design of new efflux pump inhibitors are presented.

Tuberculosis is a major health problem in the developing countries and second only to Acquired Immuno Deficiency Syndrome (AIDS) as the notorious killer worldwide due to a single infectious agent. Chapter 8 ("Integrated *in Silico* Methods for the Design and Optimization of Novel Drug Candidates: A Case Study on Fluoroquinolones – *Mycobacterium Tuberculosis* DNA Gyrase Inhibitors"), contributed by Nikola Minovski and Marjana Novič, provides 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 antibacterials as potential novel *Mycobacterium tuberculosis* DNA gyrase inhibitors. The authors start the chapter by presenting a classical industrial drug discovery pipeline followed by a brief mention

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

of different *in silico* drug design strategies. Then the authors present a QSAR-aided integrated *in silico* approach for small-scale ligand screening and virtual combinatorial library design. The authors then present a novel and simple semi-automated Boolean-based clustering method for hit(s) identification. To demonstrate the practical implementation of the suggested integrated *in silico* screening protocol, the authors present a specific example related to the design and identification of novel 6-fluoroquinolone antibacterials as potential inhibitors against *M. tuberculosis* DNA gyrase enzyme.

Hepatitis C, a chronic blood-borne virus infection, is a global concern. It is likely that we will see more mortality from this disease in the coming years. Nonstructural 5B (NS5B) polymerase and nonstructural 3/4A (NS3/4a) protease have been found to be promising targets for the development of anti-HCV (Hepatitis C Virus) agents. In Chapter 9 ("Computational Approaches for the Discovery of Novel Hepatitis C Virus NS3/4a and NS5B Inhibitors"), authored by Khac-Minh Thai, Quoc-Hiep Dong, Thi-Thanh-Lan Nguyen, Duy-Phong Le, Minh-Tri Le and Thanh-Dao Tran, different computational approaches including QSAR and virtual screening for anti-HCV drug discovery are discussed. At the beginning, the authors describe HCV genome, HCV structural and nonstructural proteins. Then the authors briefly introduce various computational modeling techniques, including QSAR, pharmacophore, and docking analyses. The authors then review different applications of QSAR and virtual screening approaches for developing NS5B polymerase inhibitors and NS3/4a protease inhibitors. The QSAR approach in the discovery for multi-target inhibitors of HCV is also described.

Alzheimer's disease is the most common form of dementia, with chronic, irreversible, and progressive neurodegenerative disorder. The acetylcholinesterase enzyme is an important target for development of drugs for the treatment of Alzheimer's disease. Chapter 10 ("QSAR Models towards Cholinesterase Inhibitors for the Treatment of Alzheimer's Disease"), authored by C Gopi Mohan and Sikhar Gupta, presents an overview of different QSAR studies carried out for a variety of cholinesterase inhibitors for the treatment of Alzheimer's disease. In the introduction, the authors discuss Alzheimer's disease and its etiologies, and then they discuss the structural aspects of the acetylcholinesterase enzyme followed by an account of different acetylcholinesterase inhibitors. A brief account of theories of QSAR, descriptors, different linear and non-linear modeling methods is then given. Finally, QSAR reports on acetylcholinesterase inhibitors for development of anti-Alzheimer's drugs are reviewed.

Cancer is a group of diseases that presents a widespread fear in society. There are several forms of cancer that can develop from almost any type of cell in the body. Breast carcinoma is a common form of cancer in females, accounting for about 20% of all cancer-related deaths. Being overexpressed in breast cancer and responsible for estrogen production, aromatase is a potential target for designing anti-breast cancer drugs. Chapter 11 ("Ligand- and Structure-Based Drug Design of Non-Steroidal Aromatase Inhibitors [NSAIs] in Breast Cancer"), authored by Tarun Jha, Nilanajn Adhikari, Amit Kumar Halder, Chanchal Mondal, and Achintya Saha, discusses the design of nonsteroidal aromatase inhibitors using QSAR and receptor-based approaches. In the Introduction of the chapter, the authors discuss breast cancer and its treatments. Then the authors highlight biochemistry, structure, and gene expression of aromatase. The authors then mention different types of aromatase inhibitors. Finally, *in silico* modeling studies of aromatase inhibitors using both ligand- and structure-based approaches are reviewed.

Predictive toxicology is an important area of application of QSAR. With any amount of resources available, it is impossible to experimentally evaluate the toxicity potential of tens of thousands of compounds against all available toxicity endpoints. Then it becomes necessary to rely on the validated computational models for risk assessment of industrial and other chemicals. Chapter 12 ("Computational Techniques

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Application in Environmental Exposure Assessment"), authored by Karolina Jagiello and Tomasz Puzyn, presents an overview of the application of computational techniques in environmental exposure assessment. In the Introduction, the authors highlight the importance of computational toxicology in the light of REACH regulations. Then the authors briefly introduce the QSAR methodology and validation tools. Multimedia mass-balance modeling and QSAR modeling for the environmental exposure assessment are described. A novel method linking multimedia mass-balance and QSAR are also commented on. The authors finally give their opinion on future possibilities of computational techniques for environmental exposure assessment.

Apart from drug discovery and predictive toxicology, the QSAR principles have also been applied in material science. Different properties and activities of various types of materials have been modeled using quantitative structure-property relationships. Nanotechnology is a newly emerging field in the 21st century. Nanomaterials have tremendous application potential in medicine, cosmetics, electronics, biomaterials, and energy production. However, the toxicity of nanomaterials to living organisms and the environment still remains relatively unknown. Thus, it is necessary to evaluate the potential adverse effects of nanomaterials to humans and the environment. Chapter 13 ("QSAR-Based Studies of Nanomaterials in the Environment"), authored by Valeria V. Kleandrova, Feng Luan, Alejandro Speck-Planche, and M. Natália D. S. Cordeiro, provides an overview of *in silico* approaches for the prediction of environmental toxicity of nanomaterials. At the beginning, the authors focus on the environmental risk of nanomaterials. Then the authors review *in silico* studies of environmental effects of nanomaterials including classical type QSAR models and perturbation models. The authors also present a case study of unified QSAR-perturbation model, demonstrating its ability to predict ecotoxicity of nanoparticles under diverse experimental conditions.

We have another contributed chapter on the computational risk assessment of nanomaterials. In Chapter 14 ("Quantitative Nanostructure-Activity Relationship Models for the Risk Assessment of NanoMaterials"), authored by Eleni Vrontakia, Thomas Mavromoustakos, Georgia Melagraki, and Antreas Afantitis, recent studies on predictive quantitative nano-structure activity relationship models for the risk assessment of nanomaterials are reviewed. The authors initially discuss the impact of nanotechnological innovations in modern society. The authors then indicate the need to evaluate the toxicity potential of nanomaterials. With a brief introduction to the basic workflow of QSAR modeling, the authors review QSAR reports on nanomaterial toxicity. The authors also give their opinion on the future of computational nanotoxicology.

QSAR practitioners have to rely on the software tools for developing validated and predictive models for possible application on new sets of data. Thus, access to validated software tools is a prerequisite for QSAR research in the present-day scenario. Most of the useful software tools are commercial, while only few tools are available free. Chapter 15 ("QSPR/QSAR Analyses by Means of the CORAL Software: Results, Challenges, Perspectives), authored by Andrey A. Toropov, Alla P. Toropova, Emilio Benfenati, Orazio Nicolotti, Angelo Carotti, Karel Nesmerak, Aleksandar M. Veselinović, Jovana B. Veselinović, Pablo R. Duchowicz, Daniel Bacelo, Eduardo A. Castro, Bakhtiyor F. Rasulev, Danuta Leszczynska, and Jerzy Leszczynski, presents a freeware tool, CORAL, for QSAR model development. The authors at first introduce some basic concepts of QSAR followed by a discussion on the technical details of CORAL. Then the authors discuss several QSAR/QSPR studies made using the CORAL software reported over the past few years. Finally, the authors indicate some problems associated with the CORAL software and mention the future directions.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

In the above 15 chapters, different aspects of theory and applications of QSAR and recent developments are discussed. This book covers the basic principles and tools used in QSAR (Chapter 1), novel descriptors (Chapters 2, 3, and 4), tests for applicability domain (Chapter 5), applications of QSARs for optimization of structures of bioactive compounds (Chapter 6), modeling of drug candidates for potential therapeutic targets in the context of drug discovery (Chapters 7-11), risk assessment of chemicals in the context of environmental safety (Chapter 12), evaluation of toxicity of nano-materials in the context of nano-science (Chapters 13 and 14), and novel software tools for application in QSAR studies (Chapter 15). I hope that this book will attract researchers in the field of medicinal chemistry and predictive toxicology both in the industry and academia.

Kunal Roy
Jadavpur University, India
November 10, 2014

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Acknowledgment

I am grateful to the members of the Editorial Advisory Board and the external reviewers for their help in developing this book. I am also thankful to Prof. L. B. Kier of Virginia Commonwealth University for writing the Foreword of this book. I also thank the authors for their contributions and the Publisher for taking an initiative for bringing out this book.

Kunal Roy
Jadavpur University, India
November 10, 2014

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 1
An Introduction to the Basic Concepts in QSAR-Aided Drug Design

Maryam Hamzeh-Mivehroud
Biotechnology Research Center & School of Pharmacy, Tabriz University of Medical Sciences, Tabriz, Iran

Babak Sokouti
Biotechnology Research Center, Tabriz University of Medical Sciences, Tabriz, Iran

Siavoush Dastmalchi
Biotechnology Research Center & School of Pharmacy, Tabriz University of Medical Sciences, Tabriz, 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.

DOI: 10.4018/978-1-4666-8136-1.ch001

Copyright © 2015, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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).

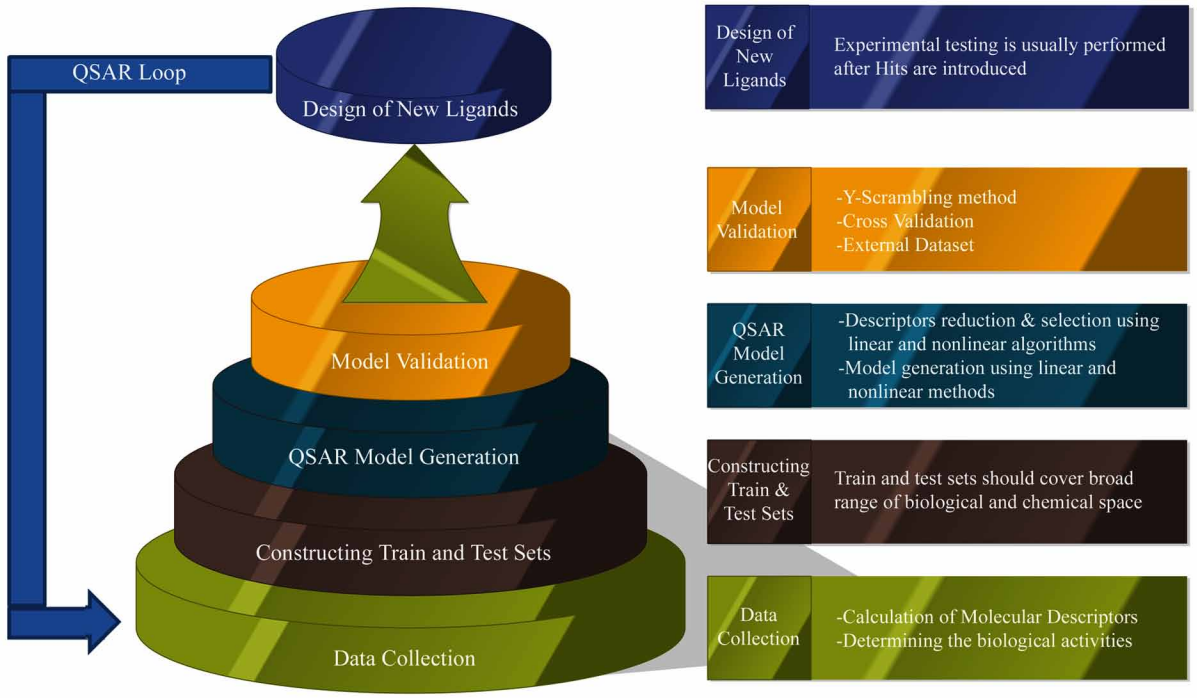
EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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, 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).

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 1. Pyramid of different steps in QSAR procedure



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 bio-

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

logical 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.

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).

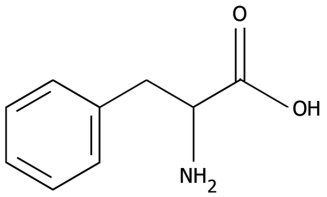
Connection tables are the other types of encoding style, in which the chemical structure is defined as a table of atom and bond types connecting the atoms with given arbitrary number in a series. Relative coordinates for each atom are usually included. An example of most commonly used connection table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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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)

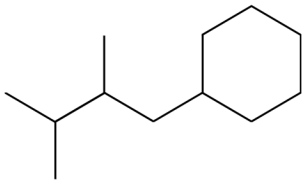
Figure 3. Molfile format for phenylalanine as an illustrative example of connection table style representation of molecules

```
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 2/14/2023 7:16 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, 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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 (v) 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 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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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.

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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:

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

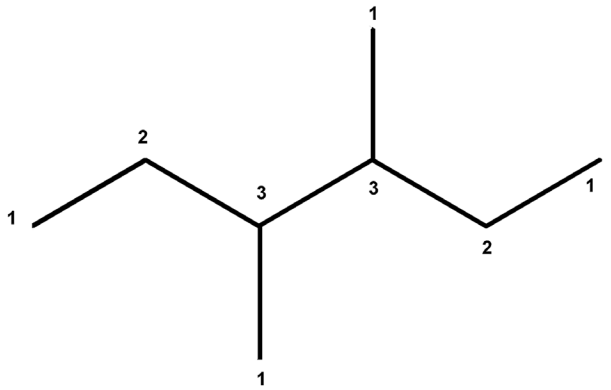
where m and n denote the degree of atoms connected by bonds. Figure 5 shows an example of a connectivity index calculation for 3,4-dimethylhexane.

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).

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

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



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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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).

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).

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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).

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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 (Babababadani & 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 outper-

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

forms 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 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).

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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).

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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.

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).

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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).

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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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
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):

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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).

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.

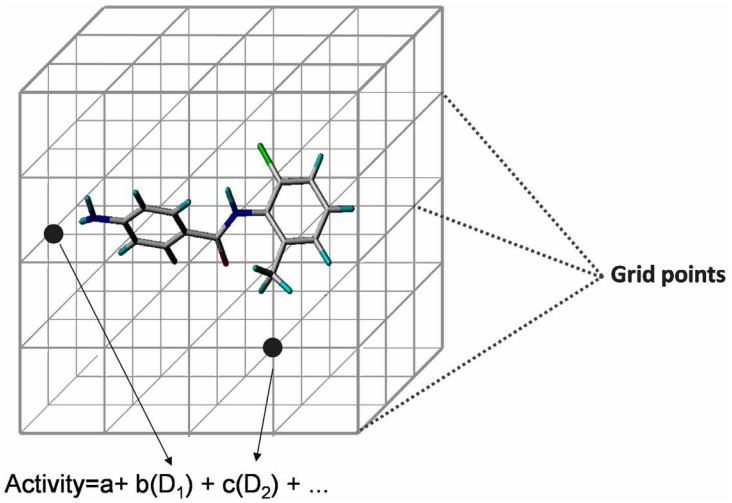
EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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.

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.



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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].

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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 & 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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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 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 multi-dimensional 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".

EBSCOhost - printed on 2/14/2023 7:16 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.

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 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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. 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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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.

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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:10.2174/13895575113136660104 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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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.

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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.

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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
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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

[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.

[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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

[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/theoecdqsartoolbox.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/.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 2
The "ETA" Indices in QSAR/QSPR/QSTR Research

Kunal Roy
Jadavpur University, India

Rudra Narayan Das
Jadavpur University, India

ABSTRACT

Descriptors are one of the most essential components of predictive Quantitative Structure-Activity/Property/Toxicity Relationship (QSAR/QSPR/QSTR) modeling analysis, as they encode chemical information of molecules in the form of quantitative numbers, which are used to develop mathematical correlation models. The quality of a predictive model not only depends on good modeling statistics, but also on the extraction of chemical features. A significant amount of research since the beginning of QSAR analysis paradigm has led to the introduction of a large number of predictor variables or descriptors. The Extended Topochemical Atom (ETA) indices, developed by the authors' group, successfully address the aspects of molecular topology, electronic information, and different types of bonded interactions, and have been extensively employed for the modeling of different types of activity/property and toxicity endpoints. This chapter provides explicit information regarding the basis, algorithm, and applicability of the ETA indices for a predictive modeling paradigm.

INTRODUCTION

Chemicals are the inseparable components of human life, and thus they can affect a wide variety of industrial, agricultural as well as household processes if not monitored properly. Elucidation of the chemistry of compounds is an essential task of the designer so as to enable easy tuning of their properties. The diverse manifestation of chemicals can be systematically explored when the study of their chemistry is suitably attended with the related biology (including toxicology), mathematics and accompanying statistics. Since the past, mathematics has been considered as an effective media of natural science (Wigner, 1960) and has been employed in different branches of science to derive logical relationship employing abstract ideas. Chemistry has a very good coherence with mathematics that leads to the quantification

DOI: 10.4018/978-1-4666-8136-1.ch002

Copyright © 2015, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

of chemical information by providing suitable platform for the generation of various algorithms. How different chemicals act differently and even the same chemical elicits different responses give the hint to nurture the 'chemistry' of compounds and explore the possible quantitative phenomena involved. The attempt to quantify the chemistry of compounds has lead to the development of several independent parameters termed as predictor variables or descriptors. These descriptors are further utilized to bring under a quantitative relationship with the responses of the chemicals commonly termed as the quantitative structure-activity/property/toxicity relationship (QSAR/QSPR/QSTR) studies. Mathematics plays a great role in deriving such predictor variables as well as in the development of correlations.

Through the journey of QSAR analysis, several predictor variables have evolved by exploring the concepts of chemistry, mathematics and physics. However, the 'graph theory' has contributed much to the development of initial ideas about descriptors. The graph theory, one of the branches in mathematics, can be traced back to Euler's Königsberg bridges problem (Euler, 1736) which led to the idea of solving problems in different fields of science namely the study of electrical circuit by Kirchhoff (Kirchhoff, 1847), chemical isomer enumeration by Cayley (Cayley, 1875) etc. In a true mathematical sense, the term 'graph' also denotes cartesian plot of data, although it was Sylvester (Sylvester, 1878) who coined the term from a contemporary chemical perspective. Application of the graph theory in chemistry leads to the concept of 'chemical graph' which was supposed to have its origin in the late eighteen century holding the hand of the Scottish chemist Cullen (Crosland, 1959). William Cullen introduced the application of 'affinity diagrams', although it remains controversial with similar type of contributions by Black (Crosland, 1959). Later Higgins (Higgins, 1789) employed the use of graphical diagrams to denote the forces between atoms in a molecule. It was Dalton (Dalton, 1808) in the nineteenth century who started the concept of ball-and-stick representation of chemical models. The idea of diagrammatic presentation of chemical compounds was very primitive during that age and through various notable discoveries the chemistry got explored gradually. The famous scientist Kekulé brought the idea of three-dimensional orientation (Fischer, 1974) and later idea of chemical valence came likewise. Considering the scope of this chapter, we would like to limit this discussion at this point.

It was in the twentieth century when the chemical graphs were explored properly, and utilizing the concepts of topological distance measures the scientific fraternity started developing theoretical variables to correlate activity of chemicals with their chemistry. Although the history of QSAR starts back in the past when Cross (Cross, 1863) reported his idea of possible correlation between chemical composition and toxicity of alcohols, the concept of theoretical descriptor did not come in action properly until the year 1947 when the Wiener index (Wiener, 1947) and the Platt number (Platt, 1947) were introduced as the first two graph theoretical descriptors for modeling boiling point of hydrocarbons. Following these milestone achievements in the realm of quantification of chemistry paradigm, a new wave was hit towards the exploration of graph theoretic parameters, and in the 1960s and 1970s the pathway was systematically imputed by the following pioneering contributions namely Simon's minimal topological difference (MTD) concept (Simon, 1974), Balaban J (Balaban, 1982), Gordon and Scantlebury index (Gordon & Scantlebury, 1964), Schultz molecular topological index (MTI) (Schultz, 1989), Hosoya Z (Hosoya, 1971), Randić's branching parameter (Randić, 1975), Kier and Hall's molecular connectivity, Kappa shape, sub-graph count, flexibility and electrotopological state (E-state) atom indices (Kier & Hall, 1986), Zagreb index by Trinajstić and co-workers (Gutman & Trinajstić, 1972), Harary index (Plavšić et al., 1993), Verloop's STERIMOL parameter (Verloop, 1987), etc. Graph theoretic descriptors are purely based on a two dimensional basis, whereas concepts of three dimension was also under parallel research and various predictor variables were also derived employing three dimensional basis. A torrential amount of research on the descriptors has led to the development of a wide variety of predictor variables till

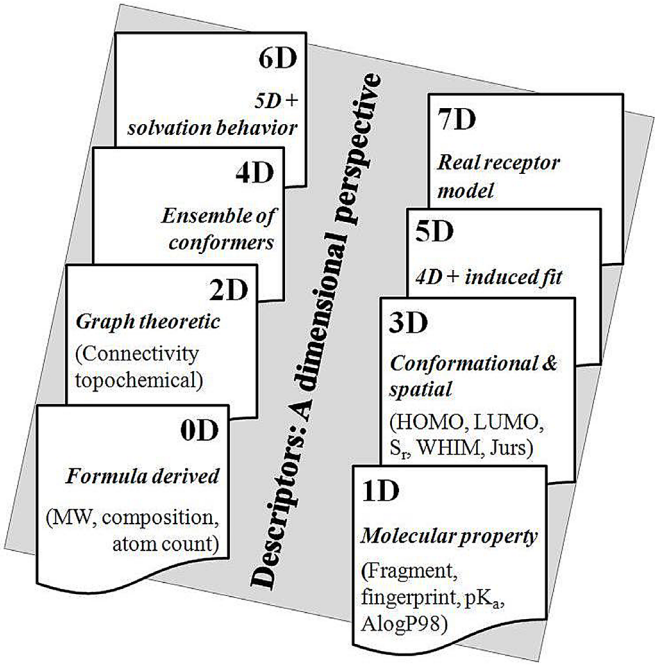
EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

today. The present chapter gives an account of such a group of two-dimensional novel descriptors named extended topochemical atom (ETA) indices which have been developed in the twenty first century from the modification of another set of topologically arrived unique (TAU) scheme descriptors of the late eighties. The ETA indices are now available under two generations and have been used in various modeling studies to predict property, activity as well as toxicity endpoints.

Descriptors in QSAR Studies: The Impact and Required Quality

Scientific deductions are based on logical components that aids in deriving conclusion from an observation. The chemoinformatics study also bears such inevitable components that ultimately help in developing predictive mathematical models and also leave a plausible explanation for the reasons involved. The molecular descriptors in QSAR studies can be viewed as the quantitative presentation of chemical structures obtained by processing the molecule using a specified mathematical algorithm. Todeschini and Consonni (Todeschini & Consonni, 2000) have ascribed descriptors as the useful numbers or the standardized experimental results obtained as the final result of a logic and mathematical procedure that leads to the transformation of chemical information which is encoded within symbolic representation of molecules. So, the mathematics is here utilized to unwrap the chemical information of compounds in the form of numerical entities. Descriptors are the algorithmic mathematical entities that strongly adhere to the chemistry of compounds and thereby form a core to the QSAR analysis. Since, the attributes of chemical compounds are depicted using different levels and dimensions of theory of chemistry and physics, descriptors are also defined on the basis of such theory coupled with a definite algorithm of mathematics. A good number of descriptors are available now-a-days corresponding to different algorithms which can be conveniently categorized from the dimensional perspective. Figure 1 shows a representative dimensional classification of various descriptors for QSAR analysis.

Figure 1. Dimensional perspective of descriptors used in predictive modeling (QSAR) analysis



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Descriptors can be defined for a whole molecule or they may also refer to specific fragments of the molecule. Again from the basis of computation, descriptors can be theoretical or experimental. Experimental descriptors correspond to various derived physicochemical properties namely partition coefficient, molar refractivity, optical rotation, pK_a values, reaction rate, parachor etc. and they usually depict characteristics of the entire molecule while theoretical descriptors are usually computationally derived mathematical entities and can define specific molecular fragment as well as the whole molecule. An interesting fact about descriptors is that they evolved from lower to higher dimensions as the practice of QSAR made its progress and so the theory. Looking back at the history of QSAR, it can be observed that Cross (Cross, 1863) was the first who observed the toxicity of primary aliphatic alcohols to be possibly correlated with their water solubility. Crum-Brown and Fraser (Crum-Brown & Fraser, 1868) are inexorably believed to be first to report the physiological action as the mathematical function of chemical constituents during an investigation of muscle paralysis activity of strychnine derivatives. Richardson (Richardson, 1869) observed a proportional relationship of narcotic effect of primary alcohols with their molecular weight and almost after two decades Richet (Richet, 1893) found the cytotoxicity of ethers, alcohols and ketones to be inversely related with their aqueous solubility measure. Then Meyer (Meyer, 1899) and Overton (Overton, 1901) came with their decisive work by correlating narcotic potency with lipid/water partition coefficient. After that, at the beginning of the twentieth century Trabe (Trabe, 1904) and Seidell (Seidell, 1912) came with their correlation analysis employing surface tension and solubility and partition coefficient respectively. Then came the pioneering discovery by Hammett (Hammett, 1935) who initially showed relationship between reaction rates with acid dissociation constant values (pK_a) and the parameter σ (sigma) which afterwards led to a "sigma-rho" culture in the field of predictive modeling analysis. Some more notable research activities were performed during this time namely correlation of narcotic activity with partition coefficient and exploring thermodynamics of chemicals by Ferguson (Ferguson, 1939), study on the involvement of ionization and shape factors towards the bacteriostatic action of aminoacridines by Albert (Albert et al., 1941) and correlation studies on antibacterial action (*E. coli*) of sulfonamides with their ionization and shape parameters by Bell and Roblin (Bell & Roblin, 1942). In this journey, it may be observed that QSAR started with the concept of correlating toxicity of chemicals with aqueous solubility by Cross whereas the term 'chemical constituents' in Crum-Brown and Fraser's hypothesis can be connected to the elemental composition of the compounds. Richardson and Richet found toxicity to be expressible in terms of molecular weight and aqueous solubility respectively. Hence, we can see that at the dawn of QSAR, scientists used different physicochemical properties, *i.e.*, experimental descriptors as variables and at the same time the concept of zero dimensional parameter like elemental composition, molecular weight were also used. The practice of using physicochemical properties like partition coefficient, solubility etc. as descriptors was also done by other scientists namely Trabe and Seidell. The concept of extracting chemical information was getting boosted during this time and Hammett led to the discovery of a theoretical parameter σ (sigma) in connection with modeling reaction kinetics. The contributions by Albert et al. and Bell & Roblin are also notable since they were the first to focus on the ionization factor and shape factors as independent parameters to derive correlation. It was in 1947, Wiener (Wiener, 1947) and Platt (Platt, 1947) came up with some more light to the path with their first topologically derived graph theoretic descriptors namely Wiener index and Platt index respectively by developing QSPR models for the boiling points of paraffin hydrocarbons. The inception of graph theoretic parameters as chemical descriptors opened a big doorway in the field

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

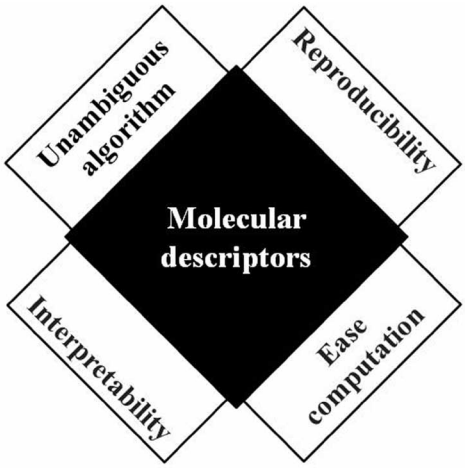
of theoretical descriptors, and since then a significant amount of research has taken place towards the development of more informative parameters. The discovery of Hammett (Hammett, 1935) was forwarded by Taft who introduced a steric parameter E_s (Taft, 1952) that distinguishes polar, steric and resonating effect exerted by chemicals and reported the first multiple regression method based linear free energy relationship (LFER) equation. In the early 1960s, several newer developments were made in the field of 2D-descriptors. In 1962 Hansch (Hansch et al., 1962) introduced a substituent hydrophobicity parameter π and later in 1964 Hansch and Fujita (Hansch & Fujita, 1964) extended the analysis by combining parameters of Hammett and Taft and developed LFER models comprising of σ, E_s and π. The analysis by Fujita and Hansch has undergone further parabolic extension and later it was also modified by Kubinyi (Kubinyi, 1976). Another concept of presenting molecular descriptors as the contribution of the parent moiety and substituents was depicted by Free-Wilson (Free & Wilson, 1964) which gave rise to 'indicator variables'. The Free-Wilson analysis was later modified by Fujita-Ban (Fujita & Ban, 1971) by incorporating logarithmic terminology.

It is to be noted that from the bases of the two-dimensional chemistry, concepts of three- dimensional (3D) molecular structure representation, electronic information and the descriptors thereof were also in the developing stage starting in 1930s. Some of the notable contributions can be cited namely the chemical bonding theory by Pauling (Pauling, 1939) and Coulson (Coulson, 1939), Sanderson's (Sanderson, 1952) electronegativity parameters, electronic distribution theory by Fukui et al. (Fukui et al., 1954) and Mulliken (Mulliken, 1955), charged partial surface area measures of Jurs & Stanton (Stanton & Jurs, 1990), shadow indices (Rohrbaugh & Jurs, 1987), WHIM (Todeschini & Lasagni, 1994) & GETAWAY descriptors by Todeschini and co-workers (Consonni et al., 2002), 3D-MoRSE descriptors (Schuur et al., 1996), EEVA descriptors, CoMFA and CoMSIA analysis by Cramer et al. (Cramer et al., 1988) and the ligand-receptor based analyses including multidimensional (4D, 5D, 6D and so on) features in the modern century (Lill, 2007). However, considering the scope of this article we would like to focus our topic on two-dimensional topological structure representation only.

Hence it is observed that the molecular descriptors of chemical compounds are pivotal as they play a very crucial role in predictive modeling analysis. Through the journey of QSAR paradigm, the attempt to generate quantitative numbers using the chemistry of compounds has yielded various mathematical descriptors. Now, the obvious question arises regarding the required quality of such essential component of QSAR analysis. It is apparently true that quantitative predictor variables were introduced to establish mathematical relationships so that predictions can be made when changes (*e.g.*, new substituents) are incorporated into structures. Since a good amount of mathematics is involved in this aspect, care should be taken such that the descriptors do not become just random numbers for developing correlation equations, rather they should encode proper chemical information. The experimental properties which are used as descriptors usually do not suffer from such problem, since they are the true reflection of the physicochemical behavior of the chemicals. However, there is possibility of incomplete information in theoretically derived parameters if they are not judged properly. Proper care is required while dealing with the mathematics since it only acts as a tool to define the final index employing several algebraic and arithmetic operations while the diagnosis of chemical information entirely belongs to the part of chemistry. Hence, two required criteria of molecular descriptors can be identified namely predictive ability and interpretability. From practical examples, in many cases it has been observed that descriptors giving high correlation against an endpoint of interest suffer from the lack of chemical diagnosis potential and vice-versa. A proper

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 2. The required features of theoretically derived descriptors



balance between this two makes a descriptor important for the QSAR analysis. This feature has also been indicated in the definition of descriptors prescribed by Todeschini and Consonni (Todeschini & Consonni, 2000) mentioned earlier as the 'useful numbers' which states the descriptors having the insight of the molecular attributes as well as ability to take part in predictive modeling analysis.

Four of the basic features of chemical descriptors are shown in Figure 2. Here, the term 'easy computation' refers to simplicity and explicitness of the descriptor algorithm since many of the higher dimensional variants require complex pretreatment operations. The descriptor should also be easily reproducible irrespective of the computational environment employed. The predictive ability refers to the skill of the descriptor in giving good correlation with an endpoint under investigation while the interpretability focuses on the chemical attributes of the analyzed molecular structures. Once a proper basis of chemical or biological activity of the molecule is established, the tuning of its features becomes easier and technically more specific. Two- dimensional descriptors are usually simple, derived using hydrogen suppressed molecular graphs with a high degree of precision and consume less computational time than the higher order indices. In the beginning of the development of graph theoretic descriptors, much more emphasis was given on the establishment of quantitative correlation than the search for mechanistic basis. As a matter of fact, many of the formerly defined descriptors were found to be flawed when chemometric modeling studies started in a much larger scale.

This article presents a group of two-dimensional descriptors, extended topochemical atom (ETA) indices, which were developed addressing several flaws of the previously described parameters. The ETA indices comprise a few graph theoretic parameters as well as some electronic indices to describe the bonding pattern, unsaturation, electronegativity and hydrogen-bonding related information. Till date, the ETA indices have been used to model several endpoints employing different categories of chemical compounds ranging from simple alkanes to complex drug molecules and agrochemicals. An account of the definition of various ETA indices with an overview of their modeling status is presented in the next sections.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

THE ETA INDICES

Brief Background

The ETA indices use information from hydrogen suppressed graphical presentation of molecules where atoms are denoted by vertices while edges represent covalent bonds. The ETA formalism was derived as an extension of another group of graph theoretic descriptors called topochemically arrived unique (TAU) scheme indices. Before getting into details of the ETA indices, we would like to provide the readers with some basic information and the motivation of developing TAU scheme parameters.

The branching parameter by Milan Randić (Randić, 1975) gave the measure of ramification of branching in a graph theoretical topological problem. This idea was extended by Kier and Hall (Kier & Hall, 1986), who gave a more comprehensive description on the type of branching by incorporating the concepts of path, cluster and chain. The molecular connectivity index of Kier and Hall's was observed to produce good correlational results with some physico-chemical and biological endpoints. However, in the late eighties Pal et al. (1988, 1989, 1990) observed several limitations in Kier and Hall's parameters and argued that predictive feature of this group of descriptors is biased, especially when higher order indices are used. The following drawbacks of Kier and Hall molecular connectivity or chi (χ) indices were identified by Pal and co-workers.

1. Kier and Hall's molecular connectivity indices chiefly portray additive information with no consideration of constitutive aspects.
2. It bears a good correlation with some quantum mechanical indices and mass spectroscopic fragmentation pattern and thereby possesses a similar pattern like logPo/w.
3. The molecular connectivity index formalism does not include information for steric interaction, stereoisomerism, resonance and aromaticity.
4. The vertex valency term δ^v often shows redundancy and becomes inconsistent when atoms of different quantum levels are considered.
5. Unlike vertex valency, the relative weight of the edges is not properly distinguished which leads to the fact that only ${}^1\chi$ and ${}^1\chi^v$ have graph theoretical meaning. Furthermore, higher order indices, i.e., ${}^m\chi^v$ (m>1), were found to possess good correlation when ${}^1\chi^v$ showed similar result and reflect that ${}^m\chi^v$ is strongly correlated with ${}^1\chi^v$.
6. The sigma and pi bonds in a system are equally weighted in this scheme.
7. Suitable adjustment in the molecular structure of several compounds of varied structures which do not possess optimum biological response in the dataset can yield the ideal value of ${}^m\chi^v$ obtained from the regression equation.

With the aim of circumventing such undesirable shortcomings of molecular connectivity indices, Pal et al. (1988, 1989, 1990) proposed a new set of topological descriptors namely TAU indices emphasizing the nuclear and electronic features of the atoms. In the TAU formalism, the vertex in a connected molecular graph has been considered to be composed of a core and a valence electronic environment. The valence electronic environment has been further partitioned into a localized and a mobile electron measure. Table 1 shows the formal definitions of various TAU parameters. It should be noted that other

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 1. Definition of the descriptors under the topochemically arrived unique (TAU) scheme formalism

Sl. No.	Type	Mathematical Definition		Note (If Any)
Parameter for the Vertex Core				
1	Core count	$\text{Core count} = \lambda = \frac{Z - Z^v}{Z^v}$ where Z=atomic number, Z^v=number of valence electron, *i.e.*, number of positive charges not neutralized by core electrons.		λ shows a regular change with Pauling's electronegativity (ε^0) across the period and along the group in an opposite manner. $1/\lambda$ gives a rough measure of the strength of the positive field of the atomic core.
Parameters for the Localized and Mobile Valence Electron Environment				
		Valence Electron Local (VEL)	**Valence Electron Mobile (VEM)**	
2	Valence electron count	$\theta' = 2h + 1.5\nu + 2l$ h is the number of bonded H-atoms, ν is the number of sigma bonds other than with hydrogen and l is the lone pairs of electrons.	$\theta = \left(8 - \theta'\right) = 8 - \left(2h + 1.5\nu + 2l\right)$	θ' gives a count of the relatively localized valence electron that the positive field of the atomic core enjoys while θ represents the count of relatively mobile electrons enjoyed by the field of the core. The number '8' in the definition of VEM denotes the number of electrons in the valence octet of a bonded atom. It was assumed that an atom sigma bonded to another non-hydrogen atom shares 50% of the other electron through the sigma bond apart from its own electrons. For a covalently linked organic atom, $h + l + \pi + \nu = 4$, π being the number of pi-bonds. So, $\theta' = 8 - \left(0.5\nu + 2\pi\right)$ and $\theta = 0.5\nu + 2\pi$. Isosteric group or isovertices are characterized by similar θ or θ′ values.
3	Vertex valency	$V_i' = \frac{\lambda_i}{\theta_i'}$	$V_i = \frac{\lambda_i}{\theta_i}$	
4	Edge weight	$E_{ij}' = E_{ji}' = \left(V_i' \times V_j'\right)^{\frac{1}{2}}$ when $i \neq j$ and i, j connected $= 0$ otherwise	$E_{ij} = E_{ji} = \left(V_i \times V_j\right)^{\frac{1}{2}}$ when $i \neq j$ and i, j connected $= 0$ otherwise	
5	Molecular index (1st order)	$T' = 0.5\sum E_{ij}'$	$T = 0.5\sum E_{ij}$	
6	Molecular index (Zero order)	${}^0T' = 0.5\sum V_i'$	${}^0T = 0.5\sum V_i$	
7	The composite index	$T = \sum_{i<j} E_{ij} = \sum_{i<j} \left(V_i \times V_j\right)^{0.5}$		
Parameter for Branching and Functional Group				
8	Functionality parameter	$F = T_R - T$ T_R being the first order VEM molecular index of the corresponding reference alkane.		A reference alkane of a molecule corresponds to a structure where all heteroatoms are replaced with carbon atom and multiple bond (covalent) with single bond. The branching parameter and the functionality parameter can be correlated as: $T = T_R - F = T_N - B - F$
9	Branching parameter	$B = T_N - T_R$ T_N being the first order VEM molecular index of the normal alkane.		

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 2. A representative list of endpoints modeled using TAU scheme indices

Endpoints Addressed	Type of Chemicals Used	Ref.
Narcosis of barnacle larvae in tomato plant and red spider	Vapor toxicity of alkanols	Pal et al., 1988
Inhibitory activity of *Mycobacterium tuberculosis*	Substituted bromophenols	Pal et al., 1989
Amoebicidal action against *Entamoeba histolytica*	Polymethylene primary diamines	Pal et al., 1990
Fuel property in terms of research octane number (RON) and performance number (PN)	Acyclic alkane.	Pal et al., 1992
Narcosis to tadpoles	Diverse functional acyclic compounds	Roy et al., 1999
General anesthetic activity	Aliphatic hydrocarbons, halocarbons and ethers	Roy et al., 2001
Heat of formation and heat of atomization	Diverse functional acyclic compounds	Roy & Saha, 2003a
Lipid–Water Partition Coefficient	Diverse functional acyclic compounds	Roy & Saha, 2003b
Aqueous solubility	Diverse functional acyclic compounds	Roy & Saha, 2003c
Molar refractivity	Diverse functional acyclic compounds	Roy & Saha, 2005

than characterizing the core and the valence electronic environment (both local and mobile), parameters defining the branching and contribution of functional groups were also defined in the TAU scheme. Several modeling studies were performed using TAU indices and reliable results were obtained. Later the same group of authors also attempted to develop models employing other descriptors along with TAU parameters so as to judge the relative contribution of TAU scheme indices in predictive modeling studies. Table 2 summarizes the various endpoints modeled employing TAU scheme parameters. It was shown by Pal et al. (1988, 1989, 1990) that isosteric groups or isovertices have similar θ or θ' values.

The Motivation for Developing ETA Indices

Nearly a decade after the development of TAU indices, Roy and Ghosh (2003) observed that even the TAU scheme has some drawbacks and there are opportunities of their further development into more appropriate descriptors. The present authors' group observed the following aspects of the TAU scheme to be inconsistent:

1. The core count 'λ' was observed to have a disproportionate incremental relationship with the atomic number of different vertices.
2. The VEM vertex count was not suitably distributed and a vertex forming sigma bond with another vertex of similar electronegativity and different electronegativity were given the same value.
3. Although it provides a weighting scheme for sigma and pi bonds, the TAU scheme does not differentiate the double bonds of different types *i.e.*, ordinary alkenes, ordinary alkynes, conjugated system, aromatic system, etc.
4. Furthermore, the TAU scheme considers the contribution of functional groups at the molecular level, which should rather be considered at the atomic level or fragmental basis.

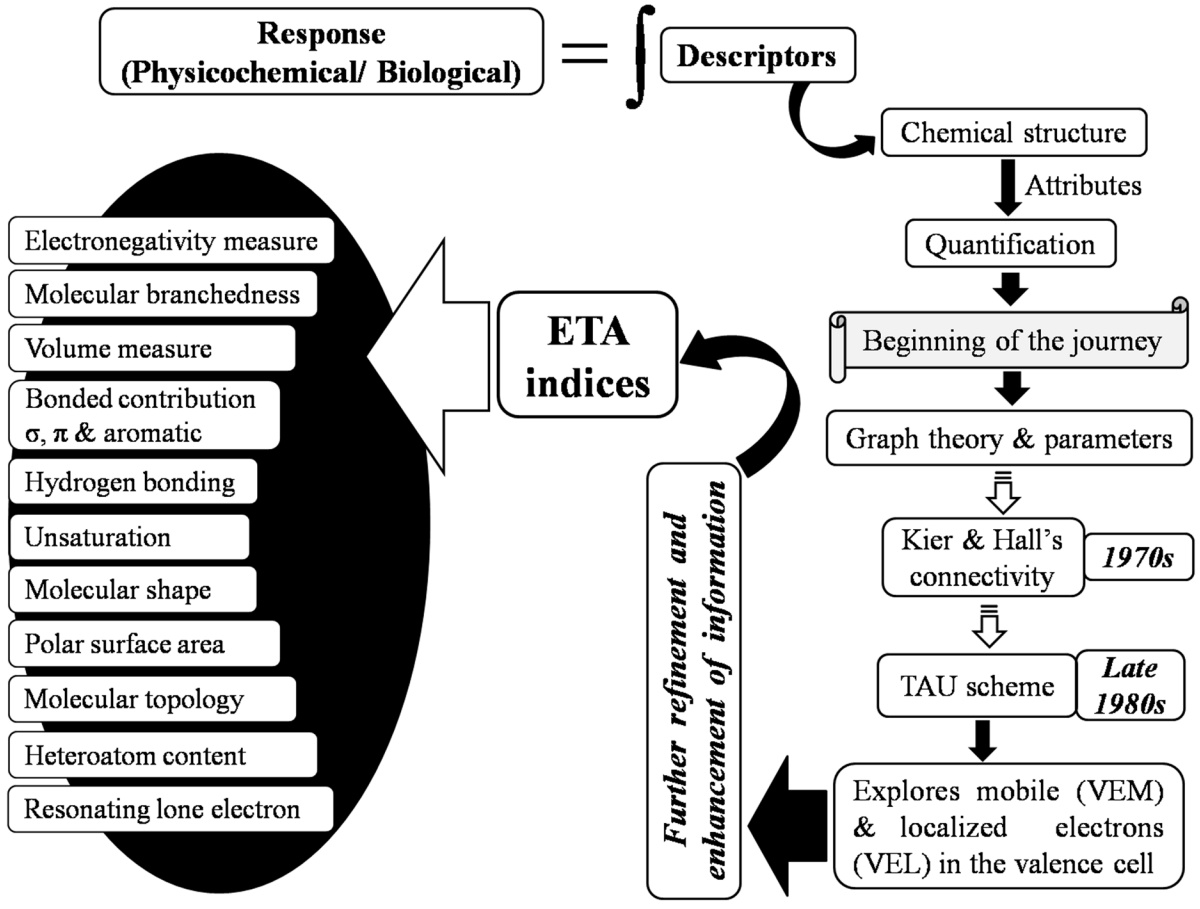
EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

These were the chief motivations to the present authors' group towards the development of a new group of predictor variables in order to achieve better diagnostic potential. In 2003, Roy and Ghosh presented a new formalism named ETA indices (Roy & Ghosh, 2003) comprising several molecular level as well as atomic level parameters. Then these indices were subjected to a thorough chemometric modeling employing different endpoints. Later in 2011, the present authors' group realized the need for some more parameters to encode the issues related to hydrogen bonding propensity of molecules, measure of polar surface area, unsaturation and electronegativity, and they came up with a second generation indices (Roy & Das, 2011). Figure 3 presents the basic features of the ETA indices from a perspective of their developmental stages.

The ETA Indices: Formalism

Presently the ETA indices comprise a set of two generation descriptors defining chemical features at the atomic as well as molecular level. Table 3 gives a generation-wise listing of the derived ETA indices. In the present article, we would present all the ETA parameters from the viewpoint of atomic and molecular basis.

Figure 3. The principle features of ETA indices along with the path of their journey



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 3. Generation-wise presentation of the ETA indices

Generation	Indices
First generation	$\sum\alpha$, $\sum\alpha/N_v$, $(\sum\alpha)_p/\sum\alpha$, $(\sum\alpha)_Y/\sum\alpha$, $(\sum\alpha)_X/\sum\alpha$, $\sum\varepsilon$, $\sum\varepsilon/N$, $\sum\beta$, $\sum\beta'$, $\sum\beta_s$, $\sum\beta'_s$, $\sum\beta_{ns}$, $\sum\beta'_{ns}$, η, η_R, η_F, η'_F, η^{local}, η_R^{local}, η_F^{local}, η'^{local}_F, η_B, η'_B
Second generation	$\Delta\alpha_A$, $\Delta\alpha_B$, ε_2, ε_3, ε_4, ε_5, $\Delta\varepsilon_A$, $\Delta\varepsilon_B$, $\Delta\varepsilon_C$, $\Delta\varepsilon_D$, ψ_1, $\Delta\psi_A$, $\Delta\psi_B$, $\Delta\beta$, $\Delta\beta'$, $\sum\beta_{ns(\delta)}$, $\sum\beta'_{ns(\delta)}$

Definition of the Atomic Level ETA Indices

The Core Count (α)

The ETA core count is formally defined as:

$$\alpha = \frac{Z - Z^v}{Z^v} \cdot \frac{1}{PN - 1} . \tag{1}$$

Here *PN* stands for period number, while Z and Z^v represent atomic number and valence electron number respectively. Hydrogen atom being considered as the reference, α for hydrogen is taken to be zero. It was observed that α values of different atoms in common organic compounds show a high correlation (R=0.946) (Roy & Ghosh, 2003) with uncorrected van der Waals volume (Moriguchi et al., 1976).

Hence, a summed value of the parameter, i.e., $\Sigma\alpha$ for all non-hydrogen atoms in a molecule gives a measure of molecular bulk. In order to incorporate relativeness to this parameter with respect to the number of vertices, another term $\Sigma\alpha/N_v$ is commonly used where N_v is the number of non-hydrogen atoms, *i.e.*, vertices in the molecule.

The Electronegativity Count (ε)

The ETA electronegativity count can be defined as follows:

$$\varepsilon = -\alpha + 0.3 \times Z^V \tag{2}$$

The term ε is characterized by a significant correlation (R=0.937) with Pauling's electronegativity scale (EN) (Roy & Ghosh, 2003).

The Hydrogen-Bonding Propensity (ψ)

The ETA hydrogen bonding propensity measure can be defines as follows:

$$\psi = \frac{\alpha}{\varepsilon} \tag{3}$$

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Based on the values of α and ε, this parameter (ψ) was defined which performs a discrimination among the atoms involved and not involved in hydrogen bonding. Table 4 gives the α, ε, and ψ values of some atoms commonly found in organic molecules.

The VEM Count (β)

In the ETA scheme, a new formalism has been incorporated to describe the VEM vertex measure. The VEM count β for a non-hydrogen vertex can be defined as:

$$\beta = \Sigma x\sigma + \Sigma y\pi + \delta \tag{4}$$

where the parameters σ and π correspond to the number of sigma and pi bonds, respectively, in the connected hydrogen suppressed molecular graph and δ represents a correction factor having a value of 0.5 per atom with loan pair of electrons capable of making resonance with an aromatic ring (e.g. nitrogen of aniline, oxygen of phenol etc.). Here x and y are the respective contributions of sigma (σ) and pi (π) bonds. We have assigned different contribution values for the sigma as well as pi bonds in the VEM count (β) of the ETA scheme. For the computation of β, contribution of a sigma bond (x) between two atoms of similar electronegativity ($\Delta\varepsilon \leq 0.3$) is considered to be 0.5, and for a sigma bond between two atoms of different electronegativity ($\Delta\varepsilon$ >0.3), it is considered to be 0.75. Again, the contributions for π bonds (y) are assigned depending on the type of double bonds as depicted below.

1. For a π bond between two atoms of similar electronegativity ($\Delta\varepsilon \leq 0.3$), the value of y is 1.
2. For a π bond between two atoms of different electronegativity ($\Delta\varepsilon$ >0.3) or for conjugated (non-aromatic) π system, y is considered to be 1.5.
3. For an aromatic π system, the value of the contribution factor is 2.

Table 4. α, ε, and ψ values of the commonly occurring atoms

Sl. No.	Atoms	α	ε	ψ
1	H	0.000	0.300	0.000
2	C	0.500	0.700	0.714
3	N	0.400	1.100	0.364
4	O	0.333	1.467	0.227
5	F	0.286	1.814	0.158
6	P	1.000	0.500	2.000
7	S	0.833	0.967	0.861
8	Cl	0.714	1.386	0.515
9	Br	1.333	0.767	1.738
10	I	1.643	0.457	3.595

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Hence, it is observable that the TAU scheme considered a value of 2.0 for all kinds of π bonds which is not in the case of ETA formalism. Furthermore, the VEM sigma contribution for a non-hydrogen vertex can be designated as β_s and defined as:

$$\beta_s = \Sigma x\sigma \tag{4a}$$

Similarly, the VEM non-sigma contribution of a non-hydrogen vertex can be designated as β_{ns} and defined as:

$$\beta_{ns} = \Sigma y\pi + \delta \tag{4b}$$

It should be noted that,

$$\beta = \beta_s + \beta_{ns}\,. \tag{4c}$$

The VEM Vertex Count (γ)

The VEM vertex count γ_i of the i^{th} vertex in a connected molecular graph can be defined as:

$$\gamma_i = \frac{\alpha_i}{\beta_i} \tag{5}$$

where α_i stands for the alpha value for the i^{th} vertex and β_i being the VEM count considering all bonds connected to the atom and lone pair of electrons (if any).

The Atom Level Composite Index ($[\eta]_i$)

The contribution of a particular vertex or position (within the common substructure in a congeneric series) to the composite index η can be determined in the following manner:

$$[\eta]_i = \sum_{j \neq i} \left[\frac{\gamma_i \gamma_j}{r_{ij}^2} \right]^{0.5} \tag{6}$$

Here, $[\eta]_i$ denotes the contribution of the i^{th} vertex to η. The term in the denominator r_{ij} corresponds to the topological distance between the i^{th} and j^{th} atoms.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Definition of the Molecular Level ETA Indices

The molecular level indices indicate computation of those descriptors which involve the entire molecular structure. There are some ETA indices which can be computed at the atomic level as well as at the whole molecular level, while there are some indices for whole molecule only. Table 5 presents a comprehensive assembly of ETA indices for the whole molecule/ molecular fragment/ substructure with their notation, mathematical definition and significance. We would just like to clarify that throughout the Table 5, N indicates total number of atoms including hydrogens while N_v corresponds to the number of vertices, *i.e.*, only non-hydrogen atoms. Again few terms are associated with a subscript 'R', e.g., N_R, $[\Sigma\alpha]_R$, η_R etc. corresponding to a reference alkane which refers to a molecular graph of the original structure where all the heteroatoms are replaced with carbon and multiple (double and triple) bonds with single bonds. Furthermore, four terms in Table 5 namely $\Delta\alpha_A$, $\Delta\alpha_B$, $\Delta\psi_A$ and $\Delta\psi_B$ are spline terms (denoted by angle bracket or chevrons, < >) meaning that any negative value will be considered zero. Regarding the computation of VEM β count, one more subtle point to be noted is that the index β is computed for edges taking the value of vertices. Since an edge is comprised of two vertices, the following relationships exist:

$\sum\beta_s$ = Sum of β_s values of all non-hydrogen vertices / 2

$\sum\beta_{ns}$ = (Sum of β_{ns} values of all non-hydrogen vertices – Sum of all $\beta_{ns(\delta)}$ values of all vertices) / 2 + Sum of all $\beta_{ns(\delta)}$ values of all vertices

$\sum\beta_{ns(\delta)}$ = Sum of all $\beta_{ns(\delta)}$ values of all vertices

Table 5. Definition of the ETA indices at the whole molecule/fragment/substructure level

Sl. No.	Mathematical Definition	Significance
1	$\sum\alpha$	Summed α values of all non-hydrogen vertices of a molecule (or fragment/ substructure). Gives a measure of molecular bulk (or for the fragment/ substructure).
2	$\Sigma\alpha/N_v$	A relative measure of molecular bulk with respect to the number of vertices.
3	$\left(\Sigma\alpha\right)_p/\Sigma\alpha$	$(\sum\alpha)_p$ stands for summation of α values of the vertices that are joined to one other non-hydrogen vertex in the connected molecular graph. Gives a measure of molecular shape.
4	$\left(\Sigma\alpha\right)_Y/\Sigma\alpha$	$(\sum\alpha)_Y$ stands for summation of α values of the vertices that are joined to three other non-hydrogen vertices in the connected molecular graph. Gives a measure of molecular shape.
5	$\left(\Sigma\alpha\right)_X/\Sigma\alpha$	$(\sum\alpha)_X$ stands for summation of α values of the vertices that are joined to four other non-hydrogen vertices in the connected molecular graph. Gives a measure of molecular shape.
6	$\Delta\alpha_A = \left\langle \frac{\sum\alpha - \left[\sum\alpha\right]_R}{N_v} \right\rangle$	A measure of count of heteroatoms.
7	$\Delta\alpha_B = \left\langle \frac{\left[\sum\alpha\right]_R - \sum\alpha}{N_v} \right\rangle$	A measure of count of hydrogen bond acceptor atoms and/or polar surface area.

continued on following page

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 5. Continued

Sl. No.	Mathematical Definition	Significance
8	$\varepsilon_1 = \frac{\sum \varepsilon}{N}$	Summed epsilon (ε) value in a molecule relative to the total number of atoms including hydrogen. Gives a measure of electronegative atom count.
9	$\varepsilon_2 = \frac{\sum \varepsilon_{EH}}{N_v}$	$\Sigma\varepsilon_{EH}$ stands for summed epsilon count of a molecule excluding hydrogen atoms. Gives a measure of electronegative atom count.
10	$\varepsilon_3 = \frac{\left[\sum \varepsilon\right]_R}{N_R}$	Summation of epsilon (ε) value relative to the total number of atoms including hydrogen in the connected molecular graph of the reference alkane.
11	$\varepsilon_4 = \frac{\left[\sum \varepsilon\right]_{SS}}{N_{SS}}$	Summation of epsilon (ε) value relative to the total number of atoms including hydrogen for a saturated carbon skeleton moiety of the normal molecule *i.e.*, carbon-carbon multiple bonds considered as single bond.
12	$\varepsilon_5 = \frac{\sum \varepsilon_{EH} + \sum \varepsilon_{XH}}{N_v + N_{XH}}$	Here, $\Sigma\varepsilon_{XH}$ stands for summed epsilon value for those hydrogen atoms which are connected to a heteroatom (e.g., −OH, $-NH_2$, −SH etc.) and N_{XH} gives a count of such hydrogens.
13	$\Delta\varepsilon_A = \varepsilon_1 - \varepsilon_3$	A measure of contribution of unsaturation and electronegative atom count.
14	$\Delta\varepsilon_B = \varepsilon_1 - \varepsilon_4$	A measure of contribution of unsaturation.
15	$\Delta\varepsilon_C = \varepsilon_3 - \varepsilon_4$	A measure of contribution of electronegativity.
16	$\Delta\varepsilon_D = \varepsilon_2 - \varepsilon_5$	A measure of contribution of hydrogen bond donor atoms.
17	$\Sigma\beta'_s$	$\Sigma\beta'_s = \Sigma\beta_s / N_v$. Summed β values for all the sigma bonds (VEM sigma contribution) relative to the number of vertices. Gives a measure of electronegative atom count of the molecule relative to the molecular size.
18	$\Sigma\beta'_{ns}$	$\Sigma\beta'_{ns} = \Sigma\beta_{ns} / N_v$. Summed β values for all the non-sigma bonds including lone electron pairs capable of resonance if any (VEM non-sigma contribution) relative to the number of vertices. Gives a measure of electron-richness of the molecule relative to the molecular size.
19	$\Sigma\beta$	$\Sigma\beta = \Sigma\beta_s + \Sigma\beta_{ns}$. Summation of β values of all non-hydrogen vertices in a molecule.
20	$\Sigma\beta'$	$\Sigma\beta' = \Sigma\beta / N_v$. Summed β values relative to the number of vertices. Gives a measure of electronic features of the molecule relative to the molecular size.
21	$\Sigma\beta_{ns(\delta)}$	A measure of lone electrons entering into resonance with an aromatic system.
22	$\sum \beta'_{ns(\delta)} = \sum \beta_{ns(\delta)} / N_v$	A measure of lone electrons entering into resonance relative to the molecular size.
23	$\Delta\beta = \sum \beta_{ns} - \sum \beta_s$	A measure of relative unsaturation content.

continued on following page

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 5. Continued

Sl. No.	Mathematical Definition	Significance
24	$\Delta\beta' = \frac{\Delta\beta}{N_v}$	A measure of relative unsaturation content with respect to the molecular size.
25	$\psi_1 = \frac{\sum\alpha}{\left[\sum\varepsilon\right]_{EH}} = \frac{\sum\alpha/N_v}{\varepsilon_2}$	A measure of hydrogen bonding propensity of the molecules and/or polar surface area.
26	$\Delta\psi_A = \left\langle 0.714 - \psi_1 \right\rangle$	A measure of hydrogen bonding propensity.
27	$\Delta\psi_B = \left\langle \psi_1 - 0.714 \right\rangle$	A measure of hydrogen bonding propensity.
28	$\eta = \sum_{i<j}\left[\frac{\gamma_i\gamma_j}{r_{ij}^2}\right]^{0.5}$	The composite index (η). Here r_{ij} stands for the topological distance between i^{th} atom and j^{th} atom. Accounts for the global topology of the molecule/fragment/ substructure under investigation.
29	η	$\eta' = \eta/N_v$. The composite index relative to molecular size.
30	$\eta_R = \left\{\sum_{i<j}\left[\frac{\gamma_i\gamma_j}{r_{ij}^2}\right]^{0.5}\right\}_R$	The composite index for reference alkane (η_R).
31	$\eta_F = \eta_R - \eta$	The functionality index (η_F). Gives a measure of the heteroatoms and multiple bonds.
32	η'_F	$\eta'_F = \eta_F/N_v$ Functionality index relative to the molecular size.
33	$\eta^{local} = \sum_{i<j, r_{ij}=1}(\gamma_i\gamma_j)^{0.5}$	The local composite index (η^{local}). Reflects the local topology of the analyzed structure since only bonded interactions are considered (r_{ij}=1).
34	$\eta_R^{local} = \left\{\sum_{i<j, r_{ij}=1}(\gamma_i\gamma_j)^{0.5}\right\}_R$	The local composite index for the corresponding reference alkane.
	$\eta_F^{local} = \eta_R^{local} - \eta^{local}$	Local functionality contribution.
35	$\eta_F'^{local}$	$\eta_F'^{local} = \eta_F^{local}/N_v$ Local functionality contribution relative to molecular size.
36	$\eta_B = \eta_N^{local} - \eta_R^{local} + 0.086 \times N_R$	ETA branching index (η_B), where N_R stands for the number of rings in the connected molecular graph of the reference alkane structure and η_N^{local} can be defined as: $\eta_N^{local} = 1.414 + (N_v - 3) \times 0.5$ when $N_v > 3$, otherwise 0.
37	η'_B	$\eta'_B = \eta_B/N_v$. It is a measure of branching relative to molecular size.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

SAMPLE CALCULATION

The ETA indices are easily computed from the information of molecular composition and two dimensional representations. The topological aspects of the ETA formalism is enlightened using the hydrogen suppressed molecular graph based distance matrix, while parameters like shape, size, electronegativity, electron-richness, heteroatom content, unsaturation, hydrogen bonding propensity etc. involve chemical attributes of atoms and bonds. Since, such calculation does not require any spatial information or energy minimization operation, these indices for a large number of compounds can readily be computed. Table 6 shows sample calculations of two compounds namely 2-cholobenzaldehyde and 3-hydroxypyridine.

Computation of the ETA Indices

At the beginning, the ETA indices (first generations only) were computed using our in-house GW–BASIC programs namely *KRETA*1 and *KRETA*2 (GW-BASIC programs, 2003) that use distance matrix and VEM vertex count as inputs. Presently, two software platforms allow the computation of ETA indices namely Dragon (version 6) (Dragon version 6, 2010) and PaDEL-Descriptor (an open source software of the NUS, http://padel.nus.edu.sg/software/padeldescriptor/) (Yap, 2011). In Dragon (version 6), only first generation ETA indices can be computed while PaDEL-Descriptor allows the computation of all indices in this scheme. Table 7 presents the list of ETA indices corresponding to the notations used in Dragon (version 6) and PaDEL-Descriptor software.

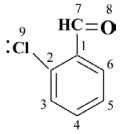
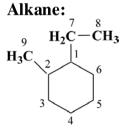
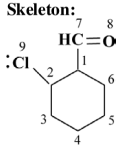
AN OVERVIEW OF THE MODELED STUDIES

The ETA indices have been subjected to extensive chemometric modeling studies in order to judge their diagnostic potential since the inception. The initial studies reflected the predictive power of the first generation indices only, while after the development of the second generation indices, the application of both generation descriptors have been exemplified. The QSAR studies employing the ETA indices cover a wide variety of chemicals ranging from simple aliphatic hydrocarbons to complex drug molecules. Recently, the ETA indices have also been successfully employed to model relatively novel chemicals like 'ionic liquids'. Table 8 gives an overview of the *in silico* models reported so far using the ETA indices. It may be noted that in all the studies, various other topological non-ETA indices have been additionally used and the ETA indices have been found to produce models of better or comparable statistical qualities in comparison to the models from non-ETA indices. In many of the cases, the quality of the developed equations improved when ETA indices have been used in combination with other non-ETA parameters. Another observation was also very interesting during the modeling of non-specific toxicity endpoints when addition of computed lipophilicity parameter AlogP to the ETA descriptor pool gave better results. It may be noted that the non-specific toxicity of chemicals is largely encoded by the ETA bulk parameter that gives a measure of lipophilicity of the molecules. The toxicity shows either a proportionate or a parabolic relationship with molecular bulk. The toxicity is also influenced by molecular branchedness as encoded by the ETA branching and shape parameters. The ETA functionality parameters additionally show the positive effect of lipophilic heteroatoms like chlorine, bromine etc. and carbonyl oxygen group towards the toxicity. The latter factor is implicated to the Schiff base formation phenomenon with biological amines. The toxicity of chemicals is found to be negatively related to the

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 6. Computation of ETA indices using two sample molecules 2-cholorobenzaldehyde and 3-hydroxypyridine

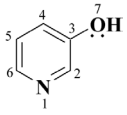
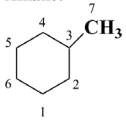
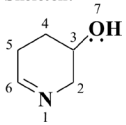
2-Cholorobenzaldehyde

Structure (Arbitrary Numbering)	Index	Atom Level Indices: Vertex (Arbitrary Numbering) 1	2	3	4	5	6	7	8	9
Normal Structure: 	α_i	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.333	0.714
	ε_i	0.700	0.700	1.000	1.000	1.000	1.000	1.000	1.467	1.386
	$[\beta_s]_i$	1.500	1.750	1.000	1.000	1.000	1.000	1.25	0.750	0.750
	$[\beta_{ns}]_i$	2.000	2.000	2.000	2.000	2.000	2.000	1.500	1.500	0.500
	β_i	3.500	3.750	3.000	3.000	3.000	3.000	2.750	2.250	1.250
	γ_i	0.143	0.133	0.167	0.167	0.167	0.167	0.182	0.148	0.571
Reference Alkane: 	$[\eta]_i$	0.875	0.886	0.783	0.720	0.718	0.777	0.757	0.519	1.036
	$[\alpha_R]_i$	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500
	$[\varepsilon_R]_i$	1.000	1.000	1.300	1.300	1.300	1.300	1.300	1.600	1.600
	$[\beta_R]_i$	1.500	1.500	1.000	1.000	1.000	1.000	1.000	0.500	0.500
	$[\gamma_R]_i$	0.333	0.333	0.500	0.500	0.500	0.500	0.500	1.000	1.000
Saturated Carbon Skeleton: 	$[\eta_R]_i$	2.272	2.260	2.226	2.092	2.110	2.250	2.264	2.169	2.353
	$[\eta'_F]_i$	0.155	0.153	0.160	0.152	0.155	0.164	0.167	0.183	0.146

Molecular Level Indices

Index	Value	Index	Value	Index	Value
$\sum\alpha$	4.547	η_R^{local}	4.343	$\Delta\varepsilon_C$	-0.119
$\sum\alpha/N_v$	0.505	η_F^{local}	2.800	$\Delta\varepsilon_D$	0.000
$(\sum\alpha)_p/\sum\alpha$	0.230	η_N^{local}	4.414	$\sum\beta'_s$	0.556
$(\sum\alpha)_Y/\sum\alpha$	0.220	η'_B	0.018	$\sum\beta'_{ns}$	0.889
$(\sum\alpha)_X/\sum\alpha$	0.000	ε_1	0.661	$\sum\beta'$	1.444
$\Delta\alpha_A$	0.005	ε_2	0.861	$\Delta\beta'$	0.333
$\Delta\alpha_B$	0.000	ε_3	0.433	$\sum\beta'_{ns(\delta)}$	0.056
η	3.536	ε_4	0.553	ψ_1	0.586
η_R	9.998	ε_5	0.861	$\Delta\psi_A$	0.128
η'_F	0.718	$\Delta\varepsilon_A$	0.228	$\Delta\psi_B$	0.000
η^{local}	1.543	$\Delta\varepsilon_B$	0.108		

3- Hydroxypyridine

Structure (Arbitrary Numbering)	Index	Atom Level Indices: Vertex (Arbitrary Numbering) 1	2	3	4	5	6	7
Normal Structure: 	α_i	0.400	0.500	0.500	0.500	0.500	0.500	0.333
	ε_i	1.100	1.000	0.700	1.000	1.000	1.000	1.767
	$[\beta_s]_i$	1.500	1.250	1.750	1.000	1.000	1.250	0.750
	$[\beta_{ns}]_i$	2.000	2.000	2.000	2.000	2.000	2.000	0.500
	β_i	3.500	3.250	3.75	3.000	3.000	3.250	1.250
	γ_i	0.114	0.154	0.133	0.167	0.167	0.154	0.266
Reference Alkane: 	$[\eta]_i$	0.500	0.587	0.665	0.627	0.594	0.548	0.574
	$[\alpha_R]_i$	0.500	0.500	0.500	0.500	0.500	0.500	0.500
	$[\varepsilon_R]_i$	1.300	1.300	1.000	1.300	1.300	1.300	1.600
	$[\beta_R]_i$	1.000	1.000	1.500	1.000	1.000	1.000	0.500
	$[\gamma_R]_i$	0.500	0.500	0.333	0.500	0.500	0.500	1.000
Saturated Carbon Skeleton: 	$[\eta_R]_i$	1.856	1.928	1.938	1.928	1.856	1.813	1.933
	$[\eta'_F]_i$	0.194	0.192	0.182	0.186	0.180	0.181	0.194

Molecular Level Indices

Index	Value	Index	Value	Index	Value
$\sum\alpha$	3.233	η_R^{local}	3.394	$\Delta\varepsilon_C$	-0.115
$\sum\alpha/N_v$	0.462	η_F^{local}	2.321	$\Delta\varepsilon_D$	0.071
$(\sum\alpha)_p/\sum\alpha$	0.103	η_N^{local}	3.414	$\sum\beta'_s$	0.607
$(\sum\alpha)_Y/\sum\alpha$	0.155	η'_B	0.015	$\sum\beta'_{ns}$	0.929
$(\sum\alpha)_X/\sum\alpha$	0.000	ε_1	0.631	$\sum\beta'$	1.536
$\Delta\alpha_A$	0.000	ε_2	0.867	$\Delta\beta'$	0.321
$\Delta\alpha_B$	0.038	ε_3	0.433	$\sum\beta'_{ns(\delta)}$	0.071
η	2.048	ε_4	0.548	ψ_1	0.533
η_R	6.267	ε_5	0.796	$\Delta\psi_A$	0.181
η'_F	0.654	$\Delta\varepsilon_A$	0.197	$\Delta\psi_B$	0.000
η^{local}	1.073	$\Delta\varepsilon_B$	0.083		

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 7. List of ETA indices corresponding to the notations used in Dragon (ver. 6) and PaDEL-Descriptor software

Dragon (Version 6) Software Platform					
Sl. No.	**Notation**	**ETA Index**	**Sl. No.**	**Notation**	**ETA Index**
1	Eta_alpha	$\sum\alpha$	13	Eta_L	η^{local}
2	Eta_alpha_A	$\sum\alpha/N_v$	14	Eta_L_A	η^{local}/N_v
3	Eta_epsi	$\sum\varepsilon$	15	Eta_F	η_F
4	Eta_epsi_A	$\sum\varepsilon/N$	16	Eta_F_A	η'_F
5	Eta_betaS	$\sum\beta_s$	17	Eta_FL	η_F^{local}
6	Eta_betaS_A	$\sum\beta'_s$	18	Eta_FL_A	η'^{local}_F
7	Eta_betaP	$\sum\beta_{ns}$	19	Eta_B	η_B
8	Eta_betaP_A	$\sum\beta'_{ns}$	20	Eta_B_A	η'_B
9	Eta_beta	$\sum\beta$	21	Eta_sh_p	$(\sum\alpha)_p/\sum\alpha$
10	Eta_beta_A	$\sum\beta'$	22	Eta_sh_y	$(\sum\alpha)_Y/\sum\alpha$
11	Eta_C	η	23	Eta_sh_x	$(\sum\alpha)_X/\sum\alpha$
12	Eta_C_A	η/N_v			
PaDEL-Descriptor Software Platform					
Sl. No.	**Notation**	**ETA Index**	**Sl. No.**	**Notation**	**ETA Index**
1	ETA_Alpha	$\sum\alpha$	22	ETA_Beta_s	$\sum\beta_s$
2	ETA_AlphaP	$\sum\alpha/N_v$	23	ETA_BetaP_s	$\sum\beta'_s$
3	ETA_dAlpha_A	$\Delta\alpha_A$	24	ETA_Beta_ns	$\sum\beta_{ns}$
4	ETA_dAlpha_B	$\Delta\alpha_B$	25	ETA_BetaP_ns	$\sum\beta'_{ns}$
5	ETA_Epsilon_1	ε_1	26	ETA_dBeta	$\Delta\beta$
6	ETA_Epsilon_2	ε_2	27	ETA_dBetaP	$\Delta\beta'$
7	ETA_Epsilon_3	ε_3	28	ETA_Beta_ns_d	$\sum\beta_{ns(\delta)}$
8	ETA_Epsilon_4	ε_4	29	ETA_BetaP_ns_d	$\sum\beta'_{ns(\delta)}$
9	ETA_Epsilon_5	ε_5	30	ETA_Eta	η
10	ETA_dEpsilon_A	$\Delta\varepsilon_A$	31	ETA_EtaP	η/N_v
11	ETA_dEpsilon_B	$\Delta\varepsilon_B$	32	ETA_Eta_R	η_R
12	ETA_dEpsilon_C	$\Delta\varepsilon_C$	33	ETA_Eta_F	η_F
13	ETA_dEpsilon_D	$\Delta\varepsilon_D$	34	ETA_EtaP_F	η'_F
14	ETA_Psi_1	ψ_1	35	ETA_Eta_L	η^{local}
15	ETA_dPsi_A	$\Delta\psi_A$	36	ETA_EtaP_L	η^{local}/N_v
16	ETA_dPsi_B	$\Delta\psi_B$	37	ETA_Eta_R_L	η_R^{local}
17	ETA_Shape_P	$(\sum\alpha)_p/\sum\alpha$	38	ETA_Eta_F_L	η_F^{local}
18	ETA_Shape_Y	$(\sum\alpha)_Y/\sum\alpha$	39	ETA_EtaP_F_L	η'^{local}_F
19	ETA_Shape_X	$(\sum\alpha)_X/\sum\alpha$	40	ETA_Eta_B_RC	η_B
20	ETA_Beta	$\sum\beta$	41	ETA_EtaP_B_RC	η'_B
21	ETA_BetaP	$\sum\beta'$			

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

*Table 8. An overview of the modeling studies performed employing ETA indices**

Sl. No.	Studied Relationship	Endpoint	Type of Chemicals	No. of Samples	Chemometric Operation	Representative Value of Metrics	Chemical Attributes Explored using ETA Indices	Ref.
1	QSTR	Toxicity towards *Tetrahymena pyriformis*	Substituted phenols	n_{total}=50	Principal component factor analysis, MLR	Q^2=0.945, R_a^2=0.950	The factors increasing the toxicity of phenols are molecular bulk, branching, η_R, η'_F, $\Sigma\beta'_{as}$, ETA functionality of phenolic O.	Roy & Ghosh, 2003
2	QSTR	Toxicity towards fish	Substituted benzenes (phenols, anilines, other hydrocarbons)	n_{total}=92	Principal component factor analysis, MLR, all-possible-subsets regression.	R^2=0.885, R_a^2=0.876, Q^2=0.865, F (df) = 92.6 (7, 84), s=0.230	The positive contribution of molecular bulk, presence of chloro, hydroxy, methyl and nitro substituents. Presence of fluoro, ether functionality, amino or nitro functionality in an otherwise unsubstituted ring, nitro group ortho to choloro reduces toxicity.	Roy & Ghosh, 2004a
3	QSTR	Toxicity towards *Tetrahymena pyriformis*	Nitroaromatic compounds	n_{total}=42	Principal component factor analysis, MLR	R^2=0.920, R_a^2=0.910, Q^2=0.880, F (df) = 101.4 (4, 37), s=0.22	Molecular bulk (size), halogen and additional nitro group in the system increases toxicity while toxicity gets decreased by the presence of methyl and hydroxymethyl substituent.	Roy & Ghosh, 2004b
4	QSTR	Acute Toxicity towards *Vibrio fischeri*	Phenylsulfonyl carboxylates	n_{total}=56	Principal component factor analysis, MLR	R^2=0.852, R_a^2=0.837, Q^2=0.726, F (df) = 57.4 (5, 50), s=0.186	The negative impact of steric bulk, branching, chloro substituent, unsaturation, electronegative atoms towards toxicity.	Roy & Ghosh, 2004c
5	QSTR	Acute Toxicity towards *Vibrio fischeri*	Phenylsulfonyl carboxylates	n_{total}=56	GFA	R^2=0.873, R_a^2=0.861, Q^2=0.771, F (df) = 69.0 (5, 50), s=0.172	The parameters reducing toxicity are steric bulk, functionality contribution, chloro group, nitro group.	Roy & Ghosh, 2005
6	QSTR	Acute Toxicity towards *Rana japonica*	Benzene derivatives	n_{total}=51	GFA, FA, MLR, PCRA	R^2=0.915, R_a^2=0.901, Q^2=0.847, F (df) = 65.841 (7, 43), s=0.183	Toxicity is parabolically related to molecular size, increases with the presence of chloro substituent, while gets reduced due to the occurrence of groups like methoxy, hydroxy, carboxy, amino.	Roy & Ghosh, 2006a

continued on following page

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 8. Continued

Sl. No.	Studied Relationship	Endpoint	Type of Chemicals	No. of Samples	Chemometric Operation	Representative Value of Metrics	Chemical Attributes Explored using ETA Indices	Ref.
7	QSTR	Nonspecific toxicity towards *Saccharomyces cerevisiae*	Substituted benzene derivatives	n_{total}=51	FA, PCA, stepwise MLR, PLS	R^2=0.884, R_a^2=0.874, Q^2=0.851, F (df) = 87.9 (4, 46), s=0.235	The negative contributions of amino & carboxylic acid functionalities on the benzene ring and occurrence of electronegative atom towards the toxicity while positive contributions of presence of chloro group and molecular branchedness.	Roy & Sanyal, 2006
8	QSPR	*n*-Octanol/ water partition coefficient	Non-ionic organic compounds	n_{total}=122; n_{train}=92, n_{test}=30	*k*-means clustering, principal component factor analysis, MLR, PLS, stepwise MLR, PCRA	Total: R^2=0.960, Q^2=0.953 Divided: R^2=0.974, Q^2=0.970, R^2_{pred}=0.906	Partion coefficient increases in the presence of molecular bulk and degree of halogen substitution while hydrogen bonding or polar interaction reduces it.	Roy et al., 2006a
9	QSPR	Bioconcentration factors in fish	Non-ionic organic compounds	n_{total}=122	Principal component factor analysis, MLR, PLS, stepwise MLR, PCRA, *k*-means clustering	R^2=0.948, R_a^2=0.944, Q^2=0.847, F (df) = 254.9 (8, 113), s=0.362	Presence of nitro, amino, hydroxyl groups reduces BCF, while it increases by branchedness, molecular volume, and chloro substituents.	Roy et al., 2006b
10	QSTR	Inhibition of seed germination rate of *Cucumis sativus*	Substituted phenols	n_{total}=41	Stepwise MLR, FA-MLR, GFA-MLR, G/ PLS, FA-PLS, PCRA	R^2=0.825, R_a^2=0.811, Q^2=0.719, PRESS=2.065	Toxicity proportionately increases in presence of branching, unsaturation and groups like nitro, cholo, bromo while functionalities namely hydroxyl, carboxy, methoxy, ortho-methyl exerts negative impacts. Furthermore, toxicity was also found to possess a non-linear relationship with ETA volume parameter $\sum\alpha$.	Roy & Ghosh, 2006b
11	QSTR	Toxicity towards *Chlorella vulgaris*	Diverse functional organic compounds	n_{total}=91, n_{train}=68, n_{test}=23	*k*-means clustering, stepwise MLR, FA-MLR, PLS, PCRA	Total: R^2=0.928, R_a^2=0.925, Q^2=0.913, F (df) = 275.29 (4, 86) Divided: R^2=0.929, Q^2=0.900, R^2_{pred}=0.870	Molecules possessing higher amount of molecular bulk ($\sum\alpha$), electron richness ($\sum\beta_{ns}$) and lipophilic substituents like chloro or bromo are more toxic.	Roy & Ghosh, 2007

continued on following page

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 8. Continued

Sl. No.	Studied Relationship	Endpoint	Type of Chemicals	No. of Samples	Chemometric Operation	Representative Value of Metrics	Chemical Attributes Explored using ETA Indices	Ref.
12	QSTR	Human toxicity data (blood/ serum conc.)	Organic compounds including pharmaceuticals	n_{total}=26, n_{train}=19, n_{test}=7	Stepwise MLR, *k*-means clustering, FA-MLR, GFA, PLS, G/PLS, PCRA	Total: R^2=0.903, R_a^2=0.895, Q^2=0.834, F (df) = 106.88 (2, 23) Divided: R^2=0.724, Q^2=0.639, R^2_{pred}=0.655	Molecular bulk, cholo group (lipophilicity), heteroatom within a chain or ring, unsaturation increase the toxicity while hydroxyl group and branching reduce it.	Roy & Ghosh, 2008
13	QSTR	Cytotoxicity in rat hepatocytes	NSAIDs	n_{total}=15	Stepwise MLR, PCRA, PLS, GFA, G/PLS	R^2=0.919, R_a^2=0.906, Q^2=0.854, F (df) = 67.92 (2, 12)	Toxicity increases with molecular bulk and degree of branching. Unsaturation and heteroatom content have additional impact on toxicity.	Roy & Ghosh, 2009a
14	QSTR	Toxicity towards *Tetrahymena pyriformis*	Diverse aromatic compounds	n_{total}=384, n_{train}=288, n_{test}=96	Stepwise MLR, *k*-means clustering, FA-MLR, PLS,	Divided: R^2=0.854, R_a^2=0.846, Q^2=0.821, R^2_{pred}=0.679	The parameters important for the prediction of toxicity are molecular bulk, halogen functionality representing lipophilicity and nitrogen containing functionality giving an account of polarity.	Roy & Ghosh, 2009b
15	QSTR	hERG K^+ channel blocking activity	Diverse functional drugs	n_{total}=67, n_{train}=50, n_{test}=17	Stepwise MLR, *k*-means clustering, FA-MLR, PLS	Divided: R^2=0.658, R_a^2=0.619, Q^2=0.540, R^2_{pred}=0.617	Electron-richness and volume increase hERG K^+ channel blocking activity while groups/fragments containing aliphatic nitrogen atom or –COOH reduce it. The parameter $\sum\alpha$ additionally showed a parabolic relationship with the modeled toxicity endpoint.	Roy & Ghosh, 2009c

continued on following page

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 8. Continued

Sl. No.	Studied Relationship	Endpoint	Type of Chemicals	No. of Samples	Chemometric Operation	Representative Value of Metrics	Chemical Attributes Explored using ETA Indices	Ref.
16	QSTR	Toxicity towards *Tetrahymena pyriformis*	Aromatic aldehydes	n_{total}=77, n_{train}=50, n_{test}=17	Stepwise MLR, GFA, G/PLS	R^2=0.895, Q^2=0.870, R^2_{pred}=0.902, $r^2_{m\ (test)}$=0.896	Molecular bulk, electronegative substituents, lipophilic substituent increase the toxicity of aromatic aldehydes. The keto oxygen in aldehyde group also imparts their effect towards toxicity.	Roy & Das, 2010
17	QSPR	Octanol-water partition coefficient	Diverse functional organic compounds	n_{total}=168, n_{train}=84, n_{test}=84	Stepwise MLR, PLS	Total (PLS): R^2=0.943, R^2_a=0.940, Q^2=0.933 Divided (PLS): R^2=0.922, Q^2=0.903, R^2_{pred}=0.919	Lipophilicity increases with molecular bulk while it tends to decrease in the presence of branching, relative unsaturation content and hydrogen bonding propensity measure.	Roy & Das, 2011a
	QSPR	Octanol-water partition coefficient	Aliphatic and aromatic compounds	n_{total}=139, n_{train}=70, n_{test}=69		Total (PLS): R^2=0.992, R^2_a=0.991, Q^2=0.989 Divided (PLS): R^2=0.988, Q^2=0.984, R^2_{pred}=0.972	Different electronic features, unsaturation and shape parameters were important for the partition coefficient of aliphatic and aromatic compounds.	Roy & Das, 2011a
	QSPR	Aqueous solubility	Aliphatic and aromatic compounds	n_{total}=193, n_{train}=97, n_{test}=96		Total (PLS): R^2=0.951, R^2_a=0.948, Q^2=0.937 Divided (PLS): R^2=0.918, Q^2=0.897, R^2_{pred}=0.872	Hydrogen bonding propensity and H-bond donor atoms were observed to be the important features for modeling aqueous solubility. However, relative unsaturation content and branchedness were also contributing.	Roy & Das, 2011a
	QSPR	Molar refractivity (R_m)	Aliphatic and aromatic compounds	n_{total}=166, n_{train}=83, n_{test}=83		Total (PLS): R^2=0.990, R^2_a=0.990, Q^2=0.989 Divided (PLS): R^2=0.980, Q^2=0.978, R^2_{pred}=0.993	Electronegativity, molecular bulk and hydrogen bonding propensity were contributing factors towards modeling of molar refraction of chemicals.	Roy & Das, 2011a
	QSPR	Aromatic substituent constant π	Aromatic substituents	n_{total}=98, n_{train}=50, n_{test}=48		Total (PLS): R^2=0.770, R^2_a=0.755, Q^2=0.718 Divided (PLS): R^2=0.795, Q^2=0.736, R^2_{pred}=0.540	Measure of H-bond acceptor atoms or polar surface area, H-bond donor atoms, H-bonding propensity, electron-richness, electronegativity, molecular bulk and branching were the contributing parameters for modeling π.	Roy & Das, 2011a

continued on following page

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 8. Continued

Sl. No.	Studied Relationship	Endpoint	Type of Chemicals	No. of Samples	Chemometric Operation	Representative Value of Metrics	Chemical Attributes Explored using ETA Indices	Ref.
	QSPR	Aromatic substituent constant MR	Aromatic substituents	n_{total}=99, n_{train}=50, n_{test}=49		Total (PLS): R^2=0.986, R_a^2=0.985, Q^2=0.979 Divided (PLS): R^2=0.977, Q^2=0.967, R^2_{pred}=0.965	MR was related to the impact of molecular bulk, relative unsaturation, electron richness, electronegativity, H-bonding atom count, H-bonding propensity, resonating lone electron pair and branching of the molecules.	Roy & Das, 2011a
	QSPR	Aromatic substituent constant σ_m	Aromatic substituents	n_{total}=97, n_{train}=49, n_{test}=48		Total (PLS): R^2=0.766, R_a^2=0.758, Q^2=0.735 Divided (PLS): R^2=0.750, Q^2=0.677, R^2_{pred}=0.736	σ_m is encoded by shape parameter, electronegativity, relative unsaturation content, resonating lone electron pair, and measure of non-hydrogen bonding bulky heteroatom.	Roy & Das, 2011a
	QSPR	Aromatic substituent constant σ_p	Aromatic substituents	n_{total}=99, n_{train}=50, n_{test}=49		Total (PLS): R^2=0.726, R_a^2=0.717, Q^2=0.682 Divided (PLS): R^2=0.670, Q^2=0.582, R^2_{pred}=0.585	σ_p can be described in terms of molecular branching, electronegativity, resonating lone electron pair, relative unsaturation content and molecular bulk.	Roy & Das, 2011a
18	QSTR	Toxicity towards *Pimephales promelas*	Diverse organic chemicals	n_{total}=459, n_{train}=344, n_{test}=115	*k*-means clustering, Stepwise MLR, GFA	Divided (GFA): R^2=0.763, Q^2=0.751, R^2_{pred}=0.783, $r^2_{m\ (test)}$=0.777	Apart from the positive impact of the computed lipophilicity measure, measures of molecular bulk, electron richness, relative unsaturation content, and branchedness are necessary attributes.	Roy & Das, 2012
19	QSPR	Critical micelle concentration (CMC)	Non-ionic surfactants	n_{total}=54, n_{train}=41, n_{test}=13	Stepwise MLR, GFA, PLS	Divided (PLS): R^2=0.907, Q^2=0.879, R^2_{pred}=0.923, $r^2_{m\ (test)}$=0.912	logCMC value increases with molecular shape, heteroatom content while molecular bulk and branching tends to reduce it.	Roy & Kabir, 2012a
21	QSPR	Critical micelle concentration (CMC)	Cationic surfactants	n_{total}=35, n_{train}=26, n_{test}=9	*k*-means clustering, stepwise MLR, GFA, ANN	Divided (GFA): R^2=0.949, Q^2=0.893, R^2_{pred}=0.904, $r^2_{m\ (test)}$=0.892	Hydrogen bond donor atoms, molecular size and degree of unsaturation affect the CMC of cationic surfactants.	Roy & Kabir, 2012b

continued on following page

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 8. Continued

Sl. No.	Studied Relationship	Endpoint	Type of Chemicals	No. of Samples	Chemometric Operation	Representative Value of Metrics	Chemical Attributes Explored using ETA Indices	Ref.
20	QSPR	Critical micelle concentration (CMC)	Anionic surfactants	n_{total}=37, n_{train}=28, n_{test}=9	Stepwise MLR, GFA, ANN, PCA	Divided (GFA): R^2=0.957, Q^2=0.938, R^2_{pred}=0.923, $r^2_{m\ (test)}$=0.823	Electronegativity and topological environment impart negative contribution, while lipophilicity and branching show positive impact towards CMC of anionic surfactants.	Roy & Kabir, 2013
22	QSPR	Aqueous solubility	Drug like molecules and agrochemicals	n_{total}=565, n_{train}=282, n_{test}=283	Stepwise MLR, GFA, G/PLS	Divided (G/PLS): R^2=0.802, Q^2=0.793, R^2_{pred}=0.806, $r^2_{m\ (test)}$=0.735	Lipophilicity, unsaturation, hydrogen bonding propensity and polar surface area measure are the responsible chemical features for logS.	Das & Roy, 2013
23	QSTR	Toxicity towards rodent and interspecies toxicity correlation among daphnids, fish and algae.	Pharmaceuticals	Rodent: n_{train}=102, n_{test}=34 Daphnia-algae: n_{train}=65, n_{test}=24 Fish-algae: n_{train}=50, n_{test}=22	Stepwise MLR, PLS	**Rodent:** R^2=0.618, Q^2=0.556, R^2_{pred}=0.550 **Daphnia-algae:** R^2=0.650, Q^2=0.580, R^2_{pred}=0.604 **Fish-algae:** R^2=0.752, Q^2=0.654, R^2_{pred}=0.696	Rodent toxicity is chiefly influenced by the charge distribution and heteroatom atom count. Interspecies correlation models show the influence of different atom-type AlogP fragments, molecular shape, charge distribution and functionality measure.	Das et al., 2013
24	QSTR	Ecotoxicity towards *Vibrio fischeri*	Diverse ionic liquids	LDA: n_{train}=110, n_{test}=37 MLR: n_{train}=90, n_{test}=36	*k*-means clustering, LDA, stepwise MLR, GFA	**LDA:** Wilk's λ=0.298, R_c=0.838, MCC_{trn}=0.841 **MLR:** R^2=0.694, Q^2=0.651, R^2_{pred}=0.739, $\overline{r^2_{m\,(\mathrm{test})}}$=0.623, $\Delta r^2_{m\ (test)}$=0.156	Branching, molecular size and solvation entropy of cations along with a lipophilicity contribution of the anions are the contributing chemical attributes for eco-toxicity of ILs towards *V. fischeri*.	Das & Roy, 2012

continued on following page

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 8. Continued

Sl. No.	Studied Relationship	Endpoint	Type of Chemicals	No. of Samples	Chemometric Operation	Representative Value of Metrics	Chemical Attributes Explored using ETA Indices	Ref.
25	QSTR	Toxicity towards *Daphnia magna*	Diverse ionic liquids	LDA: n_{train}=46, n_{test}=16 MLR: n_{train}=35, n_{test}=14	*k*-means clustering, LDA, stepwise MLR, PLS	**LDA:** Wilk's λ=0.257, R_c=0.862, MCC_{trn}=0.867, AUC-ROC=$0.991_{(trn)}$ & $1.0_{(test)}$ **MLR:** R^2=0.948, Q^2=0.875, $R^2_{\underline{pred}}$=0.817, $\overline{r^2_{m(test)}}$=0.802, $\Delta r^2_{m(test)}$=0.099	Lipophilicity, volume of heteroatoms electronegativity, presence of long cationic side chains, terminal branches and branching pattern at terminal atoms as well as at the carbon atom nearer to heteroatomic substitution are the contributing chemical features.	Roy & Das, 2013
26	QSTR	Ecotoxicity towards *Scenedesmus vacuolatus*	Diverse ionic liquids	LDA: n_{train}=42, n_{test}=18 MLR: n_{train}=35, n_{test}=14	LDA, stepwise MLR, GFA, *k*-means clustering	**LDA:** Wilk's λ=0.468, R_c=0.729, MCC_{trn}=0.802, AUC-ROC=$0.918_{(trn)}$ & $1.0_{(test)}$ **MLR:** R^2=0.883, Q^2=0.829, $Q^2_{(F1)}$=0.893, $Q^2_{\underline{(F2)}}$=0.891, $\overline{r^2_{m(test)}}$=0.796	Presence of bulky and multiple bonded heteroatom causing nucleophilic addition reaction is responsible for the toxicity. Long alkyl side chain on cationic N atom also increases toxicity. Toxicity can decrease by incorporation of H-bond donor group like hydroxyl (–OH).	Das & Roy, 2014a
27	QSTR	Inhibition of *Electrophorus electricus* AChE enzyme	Diverse ionic liquids	LDA: n_{train}=182, n_{test}=110 MLR: n_{train}=148, n_{test}=84	*k*-means clustering, LDA, stepwise MLR, GFA, PLS, molecular docking	**LDA:** Wilk's λ=0.374, R_c=0.791, MCC_{trn}=0.802, AUC-ROC=$0.959_{(trn)}$ & $0.914_{(test)}$ **MLR:** R^2=0.838, Q^2=0.808, $Q^2_{(F1)}$=0.822, $RMSE_{\underline{p}}$=0.248, $\overline{r^2_{m(test)}}$=0.762	Increased toxicity is caused by longer alkyl chain length on N atom providing proper spatial conformation, presence of quaternary N atom in ring allowing π-cationic interaction, and distributed electron density in cations due to resonating electrons. Lipophilicity of anion additionally account for the toxicity, *i.e.*, increased enzyme inhibition. However, presence of long chains on free N^+/P^+ causes reduction of the toxicity due to free rotation of chains causing binding hindrance. Toxicity also reduces due to the presence of H-bond donors & acceptors blocking partitioning of molecule.	Das & Roy, 2014b

continued on following page

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 8. Continued

Sl. No.	Studied Relationship	Endpoint	Type of Chemicals	No. of Samples	Chemometric Operation	Representative Value of Metrics	Chemical Attributes Explored using ETA Indices	Ref.
28	QSPR	Adsorption	Organic compounds	n_{total}=3483, n_{train}=2613, n_{test}=870	*k*-means clustering, PCA, stepwise MLR, PLS	Divided: R^2=0.815, Q^2=0.806, $\overline{r^2_{m(LOO)}}$=0.722, $Q^2_{(F1)}$=0.791, $\overline{r^2_{m(test)}}$=0.719	The adsorption behavior is characterized by molecular volume, H-bond donor groups, heteroatomic contribution, electronegativity, shape, H-bonding propensity and unsaturation.	Ray & Roy, 2013
29	QSPR	Bio-concentration factor	Diverse chemicals	n_{total}=522, n_{train}=324, n_{test}=198	*k*-means clustering, GFA, PLS	Divided: R^2=0.614, R_a^2=0.611, Q^2=0.597, R^2_{pred}=0.696, $\overline{r^2_{m(test)}}$=0.580	Occurrence of fused ring systems and halogen atoms exert positive contribution while electronegativity, H-bond donor groups, polarity and lowered partition coefficient have negative impact on BCF.	Pramanik & Roy, 2014
30	QSPR	Odor threshold	Aliphatic alcohols	n_{total}=53, n_{train}=42, n_{test}=11	*k*-means clustering, GFA, G/PLS	Divided: R^2=0.809, Q^2=0.778, R^2_{pred}=0.813, $\overline{r^2_{m(test)}}$=0.679	Odor threshold is potentiated by increased hydrophobicity and reduced electronegativity.	Pal et al., 2014

*Full form of some abbreviations used: **MLR:** Multiple linear regression; **GFA:** Genetic function approximation; **FA:** Factor analysis; **PCA:** Principal component analysis; **PCRA:** Principal-component regression analysis; **PLS:** Partial least squares; **LDA:** Linear discriminant analysis.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

presence of polar functional groups like hydroxy, methoxy, carboxy etc. In cases of ionic liquids, the toxicity is proportionately related to the chain length of the cationic alkyl chain. However, the same is not applicable in case of modeling AChE enzyme inhibition of eel where the head group is supposed to provide a specific spatial arrangement of the cation containing aromatic core and not for the ammonium or phosphonium core. The QSPR studies show the impact of the ETA molecular bulk parameter which relates with increased partitioning and decreased surfactancy behavior. The ETA parameters also show molecular branching and hydrogen bonding natures to be related with reduced octanol-water partitioning nature. The H-bonding phenomenon is largely attributed to the presence of polar groups like –OH, –O–, $–NH_2$, –NH– etc. However, the ETA H-bonding parameter ψ shows a negative contribution while modeling aqueous solubility of drug like compounds and agrochemicals and that has been justified by cohesive rigidity of such complex molecules due to intra- and inter-molecular H-bonding. The ETA indices for electron richness, branching and resonating lone electron pair are found to be correlated with aromatic substituent constants.

CONCLUSION AND FUTURE AVENUES

The aim of any scientific discipline lies in unraveling the reason hidden inside a system. Since, natural processes are defined by perspectives belonging to different disciplines, interdisciplinary studies are always helpful in deriving logical decisions. Predictive modeling studies aid in constituting rational basis for the physical, chemical as well as biological manifestations of chemical compounds. The chemistry suitably aided by concepts of mathematics and biology helps in unfolding the behavioral pattern of chemicals. Mathematical descriptors are the logical basis for the exploration of the chemical fraternity. Since the introduction of predictor variables, a long way has already been traversed in depicting the chemical behavior and it has been envisaged that theoretically derived chemical descriptors are more than just simple quantification of chemical structures to derive correlations. Since, there is an amalgamation of chemical principles with mathematical algorithms, care should be taken that the encoded information is also helpful in deriving mechanistic interpretation instead of developing only correlations with the endpoint under investigation.

There is also a good amount of arguments regarding the dimensionality of analysis during the development of predictive correlation models. It is quite obvious that two-dimensional analyses are relatively easier to perform and give reproducible results. However, the molecular properties and interactions are also believed to be affected by features like shape, surface area and volume which are three dimensional. Hence, it becomes very interesting if a theoretical basis is generated to incorporate information of three-dimensional perspectives in two-dimensional indices. It may be observed that there is a good correlation and inheritance between these two and three- dimensional features since the features like size, shape, volume etc. are in turn controlled by the property of the individual atoms and bonds constituting a molecule. The three-dimensional geometry is also referred to as molecular topography (Kier & Hall, 1986) and studies have shown the occurrence of quantum chemical basis (Stankevich et al., 1995) as well as Hückel's molecular orbital (Gálvez et al., 1998) features in the topological parameters rendering them reliable predictor parameters in extracting powerful chemical information.

The ETA indices are enriched since they do not rely only on topological distance based measures of hydrogen suppressed molecular graphs. This group of parameters are advantageous as they encode suitable chemical information on different aspects of molecular chemistry, *e.g.*, volume, shape, size,

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

branchedness, polarity, electronegative, electron-richness and hydrogen bonding property and also consider contribution of active hydrogen atoms. So, the ETA indices consider more than just graph theoretical aspects, employing detailed information on the atoms, bonds and the connectivity therein. Technically it will be difficult to assume any single descriptor to encode universal chemical features. Different aspects of chemical information are expressed in modeling different endpoints. In other words, endpoint like aqueous solubility or partition coefficient can be explained by molecular size, branchedness or hydrogen bonding aspects while electronic features and polarity information of a molecule may be of importance while developing predictive models against molar refractivity. The ETA formalism attempts to provide a platform of different chemical attributes so as to enable the designer with broader choice of predictor variables within a single category (Roy & Das, 2011b; Roy & Ghosh, 2010). The ETA parameters are also appended with dynamic chemical foundation suitable for deriving mechanistic basis. Moreover, all the studies performed till date using the ETA indices infer that they contain sufficient diagnostic potential for the development of predictive quantitative models for various classes of chemicals namely aliphatic compounds, aromatic compounds, agrochemicals, drugs and pharmaceuticals, ionic liquids etc. encompassing industrial, laboratory and house hold usage. Like every successful innovation walks through the development of new ideas and concepts, the ETA indices have followed a path of novel chemical exploration and we shall look forward to more nourishment of enriched chemical information in future.

REFERENCES

Albert, A., Rubbo, S. D., & Goldacre, R. (1941). Correlation of basicity and antiseptic action in an acridine series. *Nature*, *147*(3724), 332–333. doi:10.1038/147332a0

Balaban, A. T. (1982). Highly discriminating distance based topological index. *Chemical Physics Letters*, *89*(5), 399–404. doi:10.1016/0009-2614(82)80009-2

Bell, P. H., & Roblin, R. O. Jr. (1942). 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

Cayley, A. (1875). On the analytical forms called trees. *American Journal of Mathematics*, *4*(1), 266–268.

Consonni, V., Todeschini, R., & Pavan, M. (2002). Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. Part 1. Theory of the novel 3D molecular descriptors. *Journal of Chemical Information and Computer Sciences*, *42*(3), 682–692. doi:10.1021/ci015504a PMID:12086530

Coulson, C. A. (1939). The electronic structure of some polyenes and aromatic molecules. VII. Bonds of fractional order by the molecular orbital method. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, *16*(938), 413–428. doi:10.1098/rspa.1939.0006

Cramer, R. D. III, 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

Cros, A. F. A. (1863, January 9). *Action de l'alcohol amylique sur l'organisme*. Faculty of Medicine, University of Strasbourg.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Crosland, M. P. (1959). The use of diagrams as chemical 'equations' in the lecture notes of William Cullen and Joseph Black. *Annals of Science*, *15*(2), 75–90. doi:10.1080/00033795900200088

Crum-Brown, A., & Fraser, T. R. (1868). On the connection between chemical constitution and physiological action. Part I. On the physiological action of the salts of the ammonium bases, derived from strychnine, brucia, thebaia, codeia, morphia, and nicotia. *Journal of Anatomy and Physiology*, *2*(2), 224–242. PMID:17230757

Dalton, J. (1808). New system of chemical philosophy. London: Chembridge Press.

Das, R. N., & Roy, K. (2012). Development of classification and regression models for *Vibrio fischeri* toxicity of ionic liquids: Green solvents for the future. *Toxicological Reviews*, *1*, 186–195.

Das, R. N., & Roy, K. (2013). QSPR with extended topochemical atom (ETA) indices. 4. Modeling aqueous solubility of drug like molecules and agrochemicals following OECD guidelines. *Structural Chemistry*, *24*(1), 303–331. doi:10.1007/s11224-012-0080-5

Das, R. N., & Roy, K. (2014a). Predictive modeling studies for the ecotoxicity of ionic liquids towards the green algae *Scenedesmus vacuolatus. Chemosphere*, *104*, 170–176. doi:10.1016/j.chemosphere.2013.11.002 PMID:24296027

Das, R. N., & Roy, K. (2014b). Predictive *in silico* modeling of ionic liquids toward inhibition of the acetyl cholinesterase enzyme of *electrophorus electricus*: A predictive toxicology approach. *Industrial & Engineering Chemistry Research*, *53*(2), 1020–1032. doi:10.1021/ie403636q

Das, R. N., Sanderson, H., Mwambo, A. E., & Roy, K. (2013). Preliminary studies on model development for rodent toxicity and its interspecies correlation with aquatic toxicities of pharmaceuticals. *Bulletin of Environmental Contamination and Toxicology*, *90*(3), 375–381. doi:10.1007/s00128-012-0921-3 PMID:23238824

Dragon Version 6 TALETE srl, Italy. (2010). Retrieved April 30, 2014, http://www.talete.mi.it/products/dragon_description.htm

Euler, L. (1736). Solutio problematis ad geometriam situs pertinentis. *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, *8*, 128–140.

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), 387–403. doi:10.1098/rspb.1939.0030

Fischer, N. W. (1974). Kekulé and organic classification. *Ambix*, *21*(1), 29–52. doi:10.1179/amb.1974.21.1.29

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., & Ban, T. (1971). Structure-activity relation. 3. 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

Fukui, K., Yonezawa, Y., & Shingu, H. (1954). Theory of substitution in conjugated molecules. *Bulletin of the Chemical Society of Japan*, *27*(7), 423–427. doi:10.1246/bcsj.27.423

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Gálvez, J. (1998). On a topological interpretation of electronic and vibrational molecular energies. *Journal of Molecular Structure THEOCHEM*, *429*, 255–264. doi:10.1016/S0166-1280(97)00366-7

Gordon, M., & Scantlebury, G. R. (1964). Non-random polycondensation: Statistical theory of the substitution effect. *Transactions of the Faraday Society*, *60*, 604–621. doi:10.1039/tf9646000604

Gutman, I., & Trinajstić, N. (1972). Graph theory and molecular orbitals. Total π-electron energy of alternant hydrocarbons. *Chemical Physics Letters*, *17*(4), 535–538. doi:10.1016/0009-2614(72)85099-1

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). 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. 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

Higgins, W. A. (1789). *Comparative view of the phlogistic and antiphlogistic theories*. London: J. Murray.

Hosoya, H. (1971). Topological index: A newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons. *Bulletin of the Chemical Society of Japan*, *44*(9), 2332–2339. doi:10.1246/bcsj.44.2332

Kier, L. B., & Hall, L. H. (1986). *Molecular connectivity in structure-activity analysis*. Chichester, UK: Research Studies Press–John Wiley & Sons Ltd.

Kirchhoff, G. R. (1847). Über die auflösung der gleichungen, auf welche man bei der untersuchung der linearen verteilung galvanischer ströme geführt wird. *Annales der Physik et Chimie*, *72*(12), 497–508. doi:10.1002/andp.18471481202

Kubinyi, H. (1976). Quantitative structure-activity relationships, 4. Nonlinear dependence of biological activity on hydrophobic character: A new model. *Arzneimittel-Forschung. Drug Research*, *26*(11), 1991–1997. PMID:1037231

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

Meyer, H. (1899). Zur theorie der alkoholnarkose: 1. Welche eigenschaft der anästhetica bedingt ihre narkotische wirkung? *ArchArchiv for Experimentelle Pathologie und Pharmakologie*, *42*, 109–118.

Moriguchi, I., Canada, Y., & Komatsu, K. (1976). van der Waals volume and the related parameters for hydrophobicity in structure-activity studies. *Chemical & Pharmaceutical Bulletin*, *24*(8), 1799–1806. doi:10.1248/cpb.24.1799

Mulliken, R. S. (1955). Electronic population analysis on LCAO-MO molecular wave functions. I. *The Journal of Chemical Physics*, *23*(10), 1833–1840. doi:10.1063/1.1740588

Overton, C. E. (1901). *Studien uber die narkose': Zugleich ein beitrag zur allgemeinen pharmakologie*. Jena, Germany: Verlag von Gustav Fischer.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Pal, D. K., Purkayastha, S. K., Sengupta, C., & De, A. U. (1992). Quantitative structure-property relationships with TAU indices: Part I-Research octane numbers of alkane fuel molecules. *Indian Journal of Chemistry*, *31B*, 109–114.

Pal, D. K., Sengupta, C., & De, A. U. (1988). A new topochemical descriptor (TAU) in molecular connectivity concept: Part I – Aliphatic compounds. *Indian Journal of Chemistry*, *27B*, 734–739.

Pal, D. K., Sengupta, C., & De, A. U. (1989). Introduction of a novel topochemical index and exploitation of group connectivity concept to achieve predictability in QSAR and RDD. *Indian Journal of Chemistry*, *28B*, 261–267.

Pal, D. K., Sengupta, M., Sengupta, C., & De, A. U. (1990). QSAR with TAU (τ) indices: Part I – Polymethylene primary diamines as amebicidal agents. *Indian Journal of Chemistry*, *29B*, 451–454.

Pal, P., Mitra, I., & Roy, K. (2014). QSPR modeling of odor threshold of aliphatic alcohols using extended topochemical atom (ETA) indices. *Croatica Chemica Acta*, *87*(1), 29–37. doi:10.5562/cca2284

Pauling, L. (1939). *The nature of the chemical bond*. Ithaca, NY: Cornell University Press.

Platt, J. R. (1947). Influence of neighbor bonds on additive bond properties in paraffins. *Journal de Chimie Physique*, *15*, 419–420.

Plavšić, D., Nikolić, S., Trinajstić, N., & Mihalić, Z. (1993). On the Harary index for the characterization of chemical graphs. *Journal of Mathematical Chemistry*, *12*(1), 235–250. doi:10.1007/BF01164638

Pramanik, S., & Roy, K. (2014). Modeling bioconcentration factor (BCF) using mechanistically interpretable descriptors computed from open source tool "PaDEL-Descriptor". *Environmental Science and Pollution Research International*, *21*(4), 2955–2965. doi:10.1007/s11356-013-2247-z PMID:24170502

Randić, M. (1975). On characterization of molecular branching. *Journal of the American Chemical Society*, *97*(23), 6609–6615. doi:10.1021/ja00856a001

Ray, S., & Roy, K. (2013). Modeling adsorption of organic compounds on activated carbon using ETA indices. *Chemical Engineering Science*, *104*, 427–438. doi:10.1016/j.ces.2013.09.018

Richardson, B. J. (1869). Physiological research on alcohols. *The Medical Times and Gazzette*, *2*, 703–706.

Richet, C. (1893). On the relationship between the toxicity and the physical properties of substances. *Comptes Rendus Société Biologie*, *5*, 775–776.

Rohrbaugh, R. H., & Jurs, P. C. (1987). Descriptions of molecular shape applied in studies of structure/activity and structure/property relationships. *Analytica Chimica Acta*, *199*, 99–109. doi:10.1016/S0003-2670(00)82801-9

Roy, K., & Das, R. N. (2010). QSTR with extended topochemical atom (ETA) indices. 14. QSAR modeling of toxicity of aromatic aldehydes to *Tetrahymena pyriformis. Journal of Hazardous Materials*, *183*(1−3), 913–922. doi:10.1016/j.jhazmat.2010.07.116 PMID:20739120

Roy, K., & Das, R. N. (2011a). On some novel extended topochemical atom (ETA) parameters for effective encoding of chemical information and modeling of fundamental physicochemical properties. *SAR and QSAR in Environmental Research*, *22*(5−6), 451–472. doi:10.1080/1062936X.2011.569900 PMID:21598192

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Roy, K., & Das, R. N. (2011b). On extended topochemical atom (ETA) indices for QSPR studies. In E. A. Castro & A. K. Hagi (Eds.), *Advanced methods and applications in chemoinformatics: Research progress and new applications* (pp. 380–411). Hershey, PA: IGI Global.

Roy, K., & Das, R. N. (2012). QSTR with extended topochemical atom (ETA) indices. 15. Development of predictive models for toxicity of organic chemicals against fathead minnow using second generation ETA indices. *SAR and QSAR in Environmental Research*, *23*(1−2), 125–140. doi:10.1080/1062936X.2011.645872 PMID:22292780

Roy, K., & Das, R. N. (2013). QSTR with extended topochemical atom (ETA) indices. 16. Development of predictive classification and regression models for toxicity of ionic liquids towards *Daphnia magna*. *Journal of Hazardous Materials*, *254–255*, 166–178. doi:10.1016/j.jhazmat.2013.03.023 PMID:23608063

Roy, K., & Ghosh, G. (2003). Introduction of extended topochemical atom (ETA) Indices in the valence electron mobile (VEM) environment as tools for QSAR/QSPR studies. *Internet Electronic Journal of Molecular Design*, *2*(9), 599–620.

Roy, K., & Ghosh, G. (2004a). QSTR with extended topochemical atom indices. 2. Fish toxicity of substituted benzenes. *Journal of Chemical Information and Computer Sciences*, *44*(2), 559–567. doi:10.1021/ci0342066 PMID:15032536

Roy, K., & Ghosh, G. (2004b). QSTR with extended topochemical atom indices. 3. Toxicity of nitrobenzenes to *Tetrahymena pyriformis*. *QSAR & Combinatorial Science*, *23*(2–3), 99–108. doi:10.1002/qsar.200330864

Roy, K., & Ghosh, G. (2004c). 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., & Ghosh, G. (2005). QSTR with extended topochemical atom indices. Part 5. Modeling of the acute toxicity of phenylsulfonyl carboxylates to *Vibrio fischeri* using genetic function approximation. *Bioorganic & Medicinal Chemistry*, *13*(4), 1185–1194. doi:10.1016/j.bmc.2004.11.014 PMID:15670927

Roy, K., & Ghosh, G. (2006a). QSTR with extended topochemical atom (ETA) indices. VI. Acute toxicity of benzene derivatives to tadpoles (*Rana japonica*). *Journal of Molecular Modeling*, *12*(3), 306–316. doi:10.1007/s00894-005-0033-7 PMID:16249936

Roy, K., & Ghosh, G. (2006b). QSTR with extended topochemical atom (ETA) indices. 8. QSAR for the inhibition of substituted phenols on germination rate of *Cucumis sativus* using chemometric tools. *QSAR & Combinatorial Science*, *25*(10), 846–859. doi:10.1002/qsar.200510211

Roy, K., & Ghosh, G. (2007). QSTR with extended topochemical atom (ETA) indices. 9. Comparative QSAR for the toxicity of diverse functional organic compounds to *Chlorella vulgaris* using chemometric tools. *Chemosphere*, *70*(1), 1–12. doi:10.1016/j.chemosphere.2007.07.037 PMID:17765287

Roy, K., & Ghosh, G. (2008). QSTR with extended topochemical atom indices. 10. Modeling of toxicity of organic chemicals to humans using different chemometric tools. *Chemical Biology & Drug Design*, *72*(5), 383–394. doi:10.1111/j.1747-0285.2008.00712.x PMID:19012574

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Roy, K., & Ghosh, G. (2009a). QSTR with extended topochemical atom (ETA) indices. 11. Comparative QSAR of acute NSAID cytotoxicity in rat hepatocytes using chemometric tools. *Molecular Simulation*, *35*(8), 648–659. doi:10.1080/08927020902744664

Roy, K., & Ghosh, G. (2009b). QSTR with extended topochemical atom (ETA) indices. 12. QSAR for the toxicity of diverse aromatic compounds to *Tetrahymena pyriformis* using chemometric tools. *Chemosphere*, *77*(7), 999–1009. doi:10.1016/j.chemosphere.2009.07.072 PMID:19709717

Roy, K., & Ghosh, G. (2009c). QSTR with extended topochemical atom (ETA) Indices. 13. Modeling of hERG K^+ channel blocking activity of diverse functional drugs using different chemometric tools. *Molecular Simulation*, *35*(15), 1256–1268. doi:10.1080/08927020903015379

Roy, K., & Ghosh, G. (2010). Exploring QSARs with extended topochemical atom (ETA) indices for modeling chemical and drug toxicity. *Current Pharmaceutical Design*, *16*(24), 2625–2639. doi:10.2174/138161210792389270 PMID:20642426

Roy, K., & Kabir, H. (2012a). QSPR with extended topochemical atom (ETA) indices. Modeling of critical micelle concentration of non-ionic surfactants. *Chemical Engineering Science*, *73*, 86–98. doi:10.1016/j.ces.2012.01.005

Roy, K., & Kabir, H. (2012b). QSPR with extended topochemical atom (ETA) indices. 3. Modeling of critical micelle concentration of cationic surfactants. *Chemical Engineering Science*, *81*, 169–178. doi:10.1016/j.ces.2012.07.008

Roy, K., & Kabir, H. (2013). QSPR with extended topochemical atom (ETA) indices. Exploring effects of hydrophobicity, branching and electronic parameters on logCMC values of anionic surfactants. *Chemical Engineering Science*, *87*, 141–151. doi:10.1016/j.ces.2012.10.002

Roy, K., Pal, D. K., De, A. U., & Sengupta, C. (1999). Comparative QSAR studies with molecular negentropy, molecular connectivity, STIMS and TAU indices: Part I – tadpole narcosis of diverse functional acyclic compounds. *Indian Journal of Chemistry*, *38B*, 664–671.

Roy, K., Pal, D. K., De, A. U., & Sengupta, C. (2001). Comparative QSAR studies with molecular negentropy, molecular connectivity, STIMS and TAU indices: Part II- General anaesthetic activity of aliphatic hydrocarbons, halocarbons and ethers. *Indian Journal of Chemistry*, *40B*, 129–135.

Roy, K., & Saha, A. (2003a). Comparative QSPR studies with molecular connectivity, molecular negentropy and TAU indices part I: Molecular thermochemical properties of diverse functional acyclic compounds. *Journal of Molecular Modeling*, *9*(4), 259–270. doi:10.1007/s00894-003-0135-z PMID:12827454

Roy, K., & Saha, A. (2003b). Comparative QSPR studies with molecular connectivity, molecular negentropy and TAU indices. Part 2. Lipid–water partition coefficient of diverse functional acyclic compounds. *Internet Electronic Journal of Molecular Design*, *2*, 288–305.

Roy, K., & Saha, A. (2003c). QSPR with TAU indices: Water solubility of diverse functional acyclic compounds. *Internet Electronic Journal of Molecular Design*, *2*, 475–491.

Roy, K., & Saha, A. (2005). QSPR with TAU indices: Molar refractivity of diverse functional acyclic compounds. *Indian Journal of Chemistry*, *44B*, 1693–1707.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Roy, K., & Sanyal, I. (2006). QSTR with extended topochemical atom indices. 7. QSAR of substituted benzenes to *Saccharomyces cerevisiae*. *QSAR & Combinatorial Science*, *25*(4), 359–371. doi:10.1002/qsar.200530172

Roy, K., Sanyal, I., & Ghosh, G. (2006a). QSPR of *n*-octanol/water partition coefficient of nonionic organic compounds using extended topochemical atom (ETA) indices. *QSAR & Combinatorial Science*, *26*(5), 629–646. doi:10.1002/qsar.200610112

Roy, K., Sanyal, I., & Roy, P. P. (2006b). QSPR of the bioconcentration factors of non-ionic organic compounds in fish using extended topochemical atom (ETA) indices. *SAR and QSAR in Environmental Research*, *17*(6), 563–582. doi:10.1080/10629360601033499 PMID:17162387

Sanderson, R. T. (1952). Electronegativity I. Orbital electronegativity of neutral atoms. *Journal of Chemical Education*, *29*, 540–546.

Schultz, H. P. (1989). Topological organic chemistry. 1. Graph theory and topological indices of alkanes. *Journal of Chemical Information and Computer Sciences*, *29*(3), 227–228. doi:10.1021/ci00063a012

Schuur, J., Selzer, P., & Gasteiger, J. (1996). The coding of the three-dimensional structure of molecules by molecular transforms and its application to structure–spectra correlations and studies of biological activity. *Journal of Chemical Information and Computer Sciences*, *36*(2), 334–344. doi:10.1021/ci950164c

Seidell, A. (1912). A new bromine method for the determination of thymol, salicylates, and similar compounds. *Am. Chem. J.*, *47*, 508–526.

Simon, Z. (1974). Specific interactions. Intermolecular forces, steric requirements, and molecular size. *Angewandte Chemie International Edition in English*, *13*(11), 719–727. doi:10.1002/anie.197407191

Stankevich, I. V., Skovortsova, M. I., & Zefirov, N. S. (1995). On a quantum chemical interpretation of molecular connectivity indices for conjugated hydrocarbons. *Journal of Molecular Structure THEOCHEM*, *342*, 173–179. doi:10.1016/0166-1280(95)90111-6

Stanton, D. T., & Jurs, P. C. (1990). Development and use of charged partial surface area structural descriptors in computer-assisted quantitative structure–property relationship studies. *Analytical Chemistry*, *629*(21), 2323–2329. doi:10.1021/ac00220a013

Sylvester, J. J. (1878). Chemistry and algebra. *Nature*, *17*(432), 284–284. doi:10.1038/017284a0

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

Todeschini, R., & Consonni, V. (2000). *Handbook of molecular descriptors*. Weinheim, Germany: Wiley-VCH. doi:10.1002/9783527613106

Todeschini, R., Lasagni, M., & Marengo, E. (1994). New molecular descriptors for 2D- and 3Dstructures. Theory. *Journal of Chemometrics*, *8*(4), 263–273. doi:10.1002/cem.1180080405

Traube, J. (1904). Theorie der osmose and narkose. *Archiv für die gesamte Physiologie des Menschen und der Tiere, 105*(11−12), 541−558.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

Wigner, E. P. (1960). The unreasonable effectiveness of mathematics in the natural sciences. Richard courant lecture in mathematical sciences delivered at New York University, May 11, 1959. *Communications on Pure and Applied Mathematics*, *13*(1), 1–14. doi:10.1002/cpa.3160130102

Yap, C. W. (2011). PaDEL-Descriptor: An open source software to calculate molecular descriptors and fingerprints. *Journal of Computational Chemistry*, *32*(7), 1466–1474. doi:10.1002/jcc.21707 PMID:21425294

KEY TERMS AND DEFINITIONS

2D-Descriptor: A method where the considered chemical information corresponds to two-dimensional geometry of the molecule.

Chemical Graphs: A method of presenting chemical structures using a "hydrogen-suppressed" molecular skeleton where 'atoms' are denoted by 'vertices' and 'bonds' are represented by 'edges'.

Descriptor: A numerical quantity for the depiction of molecular structure using a suitable algorithm. Used as predictor variables in QSAR modeling analysis.

Extended Topochemical Atom (ETA) Indices: Indices where the topological information has been enriched using suitable electronic, heteroatom count & type, hydrogen-bonding and other essential chemical attributes. These are used as descriptors in QSAR modeling.

Hydrogen-Suppressed: A graph theoretical method of representing chemical structures where the hydrogen atoms attached to carbon and other atoms are not shown. Also termed as hydrogen-depleted.

QSAR Analysis: A method of developing mathematical correlation between a response and descriptors, *i.e.*, predictor variables for a series of chemical data.

Topology: A method assessing chemical structures using distance and connectivity information considering hydrogen-suppressed molecular graph. The distance corresponds to minimum path between vertices.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 3
Evolution of Multivariate Image Analysis in QSAR: The Case for a Neglected Disease

Matheus P. Freitas
Federal University of Lavras, Brazil

Mariene H. Duarte
Federal University of Lavras, Brazil

ABSTRACT

Multivariate Image Analysis applied in Quantitative Structure-Activity Relationship (MIA-QSAR) is a simple method to achieve, at least in a variety of examples, QSAR models with predictive abilities comparable to those of sophisticated tridimensional methodologies. MIA-QSAR is based on the correlation between properties (e.g. biological activities) and chemical descriptors, which are pixels of images representing chemical structures in a congeneric series of molecules. The MIA-QSAR approach has been improved since its creation, in 2005, both in terms of data analysis and development of more descriptive information. This chapter reports the MIA-QSAR method, including its augmented version, named aug-MIA-QSAR because of the introduction of new dimensions to better encode atomic properties. In addition, the application to a case study illustrates the main practical differences between traditional and augmented MIA-QSAR. The use of a neglected disease as example represents a challenge in QSAR, which is particularly focused on diseases with higher economical appearance.

INTRODUCTION

Quantitative structure-activity relationship (QSAR) methods have shown to be powerful tools to comprehend the action mechanisms and structural profiles required to improve a drug performance, as well as to estimate the bioactivity of a non-existing, proposed drug candidate. QSAR approaches are genuinely structure-based methods, in which biological properties are reflected by structural changes in molecules, despite the receptor information included in the modern multidimensional QSAR techniques. However, traditional QSAR, which is based on descriptors easily accessible or calculated, such as log*P*, connectiv-

DOI: 10.4018/978-1-4666-8136-1.ch003

Copyright © 2015, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

ity indices and a variety of other parameters derived from the 2D structure of a molecule, is not inferior to methods based on the three-dimensional molecular structure, at least in many practical cases (Brown & Martin, 1997; Estrada, Molina, & Perdomo-López, 2001).

Multivariate image analysis applied in QSAR (MIA-QSAR) appeared to correlate drawings of molecular structures with the corresponding biological activities (Freitas, Brown, & Martins, 2005); therefore, it is essentially a 2D QSAR technique, since descriptors are obtained from the projection of a molecule in the plane. The structural changes in a congeneric series of drug-like compounds explain the variance in the bioactivities block; in MIA-QSAR, structural changes correspond to different coordinates of the pixels composing the molecular drawings. Because the substitution pattern along with the congeneric series of compounds is captured in the calibration step, usually performed using partial least squares (PLS) regression, prediction of the bioactivities of similar compounds are often feasible. However, much chemical information is lost when a given substituent is represented as letters in the drawings, although other molecular properties, such as steric effects (*e.g.* for large side chains) and shape (*e.g.* the hexagonal benzene ring), are appropriately encoded. For instance, the MIA-QSAR model is capable of recognizing that a bromine substituent bonded to a given aromatic carbon causes an effect on the bioactivity of a molecular scaffold, but its description as "Br" in the drawing does not have chemical meaning. Thus, an augmented version for the MIA-QSAR method (aug-MIA-QSAR) was developed by introducing "dimensions" to better encode atomic sizes (using spheres with sizes proportional to the van der Waals radii) and different types of atoms (using spheres with different colors) in a molecule (Nunes & Freitas, 2013). Because aug-MIA-QSAR has been recently implemented, there are many challenges to improve its predictive ability, as well as its chemical interpretation; research directions include testing regression and variable selection methods for the aug-MIA descriptors, and also searching for ways to indicate how different atomic sizes and colors impact the trends in bioactivity in a series of drug-like compounds.

QSAR methods have been used in numerous studies in order to find a correlation between chemical structures and biological activities related to profitable diseases, such as obesity, sexual dysfunction and hypertension. Nevertheless, little attention has been devoted to neglected diseases, which mostly affect poor people from the third world, being forgotten by the big pharmaceutical companies (Ramalho, Freitas, & da Cunha, 2012). This chapter describes the development of a variety of QSAR methods during decades, focusing on the MIA-based approaches, which is demonstrated here to a set of thiosemicarbazones as anti-*Trypanosoma cruzi* agents (Garkani-Nejad & Ahmadi-Roudi, 2010). The efficacy of the current chemotherapy against *T. cruzi* is quite variable in different regions of high endemicity, because the high biological, biochemical and genetic diversity of *T. cruzy* strains (Zingales et al., 2009). Thus, the development of drugs based upon the structural optimization of existing compounds, such as the Nifuroxazide, is advantageous, considering time and efforts consumption.

Background

QSAR methods are based on the possibility of the bioactivity (or any other property) in being a function of the molecular structure, that is bioactivity = *f*(structure), in which the structure is represented by molecular parameters called descriptors. This aims at planning new substances with improved therapeutic profile. Therefore, this research field is expected to be of general interest, because the use of QSAR methods avoids exhaustive exploratory syntheses, since these methods enable designing molecules with well-defined properties, reducing time and costs during the drug development.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

The earlier studies by Cros in 1863 at the University of Strasbourg indicated that toxicity of alcohols to mammals increased as the water solubility of the alcohols decreased (Selassie, 2003). Afterwards, Crum-Brown and Fraser (1869) demonstrated the concept that molecular structure influences the biologic activity of chemical entities and that alteration in structure produces changes in biological action. Richet (1893) proposed that toxicity of some alcohols and ethers was inversely related to the corresponding water solubility and, seven years later, Meyer and Overton independently described a parameter, which can be considered the precursor of the octanol/water partition coefficient (Meyer, 1899; Overton, 1901). An important point along with this trajectory was the introduction of the Hammett constant (σ) to relate the ionization constants of *meta* and *para* substituted benzoic acids (Hammett, 1935). This finding opened the exploration of a variety of properties that correlate with biological outcomes, *e.g.* the relationship between several properties and the toxicity of homolog series of compounds (Ferguson, 1939). In 1964, Free and Wilson (1964) developed a mathematical procedure to encode bioactivities by means of the position of chemical groups in a given molecular scaffold. Simultaneously, Hansch and Fujita (1964) systematized the technique when introduced log*P* (the logarithm of the octanol/water partition coefficient) to correlate with the biological activity of a series of compounds against a variety of living organisms.

Different QSAR methods appeared since the milestone work by Hansch and Fujita (1964), but this field gained high appearance again years later, when Cramer and coworkers (1988) developed a QSAR method based on the tridimensional structure of a congeneric series of molecules. The treatment of the large amount of data generated was feasible because of the advances in QSAR technologies, improved with computers and programs. According to the so called CoMFA method (Comparative Molecular Field Analysis), each molecule of a given data set is put inside a three-dimensional lattice, whose intersections are probed by steric and electrostatic fields, generating the molecular descriptors. Contour plots representing structural moieties affecting either positively or negatively the bioactivity are given, which are possibly the main outcomes of this method. A similar approach was developed a few years later to account for the molecular similarity indices in a comparative analysis (CoMSIA); the CoMSIA maps highlight those regions within the area occupied by the ligand skeletons that require a particular physicochemical property important for activity. Hydrogen bond has been included as another descriptor in recent CoMSIA models. At least, two steps are required to start building QSAR models based on three-dimensional methods: conformational screening and three-dimensional alignment. These are not easy tasks, since the energy minimum of an optimized structure does not necessarily reflect the bioactive ligand conformation. Indeed, it has been shown for the herbicide 2,4-D that different global minima are achieved using different levels of theory for geometry optimization and conformational search, which are also different from the crystal, bioactive ligand conformation (Freitas & Ramalho, 2013). Consequently, the contour plots indicating regions with positive, neutral and negative influence on the bioactivity may not make sense, because they are dependent on the conformation, which can be either the bioactive one or not. It is worth reinforcing that bioactivity is almost always dependent on conformation and, therefore, it should be useful in QSAR studies. A recent example of this is the role of the synclinal conformation of sulfonylureas to explain the inhibition of acetohydroxyacid synthase (Jana, Delgado, & Medina, 2014); however, the correct conformation is sometimes difficult to be precisely predicted and selected.

Multidimensional (nD) QSAR methods, such as the 4D formalism in QSAR (Hopfinger et al., 1997), try to minimize the problem of choosing a single conformation for modeling, by performing ensemble averaging, the "fourth dimension". However, the bioactive conformation is not likely to be

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

represented by an average. Introduction of receptor information in the so-called receptor-dependent 4D-QSAR (RD-4D-QSAR) (Pan, Tseng, & Hopfinger, 2003) and 5D-QSAR (Vedani & Dobler, 2002) appears to better encode structural properties for further correlation with the dependent variables (the bioactivities), but the predictive abilities obtained by many nD models have not been improved significantly in comparison with traditional 3D approaches. Other dimensions have been included in QSAR to refine the molecular representation, such as multiple solvation scenarios in the 6D-QSAR, the real receptor or target-based receptor model data in the 7D-QSAR, and tautomerization (Vedani, Dobler, & Lill, 2005; Polanski, 2009; Martin, 2009). Nevertheless, classical descriptors representing connectivity indices, atom counting, molecular weight, molar refractivity, polarizability, hydrophobicity, orbital energies, etc. (which are physicochemical descriptors mostly derived from 1D and 2D molecular representation) have not shown to be inferior to 3D descriptors, because the prediction performance is often acceptable and interpretable (Brown & Martin, 1997; Estrada, Molina, & Perdomo-López, 2001).

The predictive ability of models based on these descriptors has been significantly improved with the development of mathematical methods to, for instance, select variables and to account for nonlinearities. Recent examples of feature selection methods in QSAR include ant colony optimization (ACO) and forward stepwise (FS) applied in the modeling of the anti-HIV-1 activities of 3-(3,5-dimethylbenzyl)uracil derivatives (Goodarzi, Freitas, & Jensen, 2009a), as well as genetic algorithm (GA) and successive projections algorithm (SPA) in the analysis of glycogen synthase kinase-3β inhibitory activities of a series of compounds (Goodarzi, Freitas, & Jensen, 2009b). Partial least squares (PLS) regression has been the preferred method to deal with a large amount of independent variables, while multimode methods, like parallel factor (PARAFAC) and multilinear PLS (N-PLS), have gained widespread use to deal with N-way arrays (Freitas et al., 2008). Artificial neural networks (ANN) and support vector machines (SVM) have been successfully employed to account for nonlinearities in QSAR models, improving importantly the predictive accuracy in comparison to multiple linear regression (Goodarzi, Freitas, & Jensen, 2009b).

In line with this, the multivariate image analysis applied to QSAR (MIA-QSAR) method matches the characteristics of 2D QSAR approaches, because the two-dimensional representation of molecules (molecular structure drawings) gives rise to the descriptors used in the modeling. However, 3D information, such as enantiomers, can be encoded by drawing wedge or hashed bonds at the chiral center, such as demonstrated in the MIA-QSPR modeling of the electrophoretic enantioseparation of aromatic amino acids/esters (Goodarzi & Freitas, 2009). Because nothing is expected to be more direct to represent a chemical structure than itself, such a method has provided satisfactory prediction performance in the modeling of physical, chemical and biological properties. Nevertheless, the MIA-QSAR is far from perfection, e.g. because of the lack of interpretability for descriptors representing chemical substituents, like atoms represented by their chemical symbols. Thus, the MIA-QSAR method has been constantly improved and its reliability has been attested by rigorous validation protocols, which have been continuously updated. The MIA-QSAR method and its evolution are described in the next sections and, afterwards, an application to a neglected disease is given. Trypanosomiasis was chosen because a few studies have devoted attention to this disease, which is quite common in tropical, third world countries and thousands of deaths occur worldwide, especially in the Latin America, due to the Chagas disease (caused by the parasitic protozoan *T. cruzi*).

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

MULTIVARIATE IMAGE ANALYSIS IN QSAR

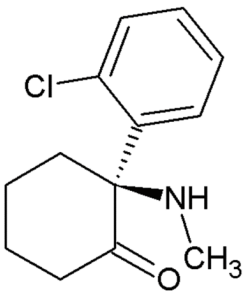
Traditional MIA-QSAR

What is the best way to describe a chemical structure, which will be correlated with its biological activity? Orbital energies, hydrophobicity, electrostatic and steric fields? According to the MIA-QSAR concept, a chemical structure is best represented by itself, as a drawing. How the structure is drawn is the main challenge to achieve better QSAR models. Other challenges include the use of different mathematical techniques for regression and variable selection, as well as the search for ways to analyze the MIA descriptors in order to indicate which pixel (or set of pixels) is responsible to increased or decreased bioactivity and why. This chapter is mainly devoted to present the development of the MIA-QSAR methodology and to show the way in representing chemical structures affects the predictive ability of the MIA-QSAR method.

Chemical structures in the traditional MIA-QSAR look like wireframes, with lines representing chemical bonds and chemical symbols to indicate atoms when necessary. Eventually, chiral centers can be designated as wedge or hashed bonds. These chemical structures can be easily drawn as routinely performed using a variety of programs, such as ChemDraw and ChemSketch. An example of a chemical structure drawn with the ChemSketch program is given in Figure 1.

QSAR methods try to search for correlations between an homolog set of compounds and the corresponding bioactivities. Thus, the set of compounds retain a structural similarity (usually the pharmacophoric group), which should be used in the 2D alignment. Such an alignment is fundamental to make variable only those moieties not common along with the set of compounds, which are responsible for the variance in the bioactivities of the molecules. The variable moieties are evident after superposition of all image files, as in the example of Figure 2. In order to superpose the files for the 2D alignment, each chemical structure should be drawn systematically, in such a way that common substructures do not vary, while substituents or variable moieties should be drawn in the same way every time it is repeated in different chemical structures. Each chemical structure should be pasted in a workspace with $x \times y$ dimension (*e.g.* 500×500 pixels size) using, for instance, the Paint application of the Microsoft Windows. Before saving the images, each one should be completely selected and a given pixel, which is common along with the set of structures, should be used to manually move (preferably using the mouse) the whole structure to a given position in the workspace. The 2D alignment is done and the image can be saved *e.g.* as bitmaps (.bmp). This procedure is pictographically represented in Figure 3. The file extension and the way in which

Figure 1. The analgesic S-ketamine drawn with the ChemSketch program



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

substituents are represented (*e.g.* CH_3 or Me to indicate a methyl group) have not shown to statistically affect the model response (Goodarzi, Freitas, & Ferreira, 2009). The drawings should be readable for further regression against the bioactivity values. Because the images are composed by pixels, these can be converted to numbers, according to the RGB (red-green-blue) system of color composition. For the chemical structure of Figure 1, the black pixels (absence of colors) are zeroed, while the blank spaces (white pixels as a sum of the red, green and blue colors, valued with 255 each) are represented by 765. The different coordinates of the pixels composing each image correspond to the chemical variance and will explain the changes in the bioactivity values. Such a numerical transformation can be performed using an appropriate program. The following scripts load and convert a bitmap image into a set of numbers, using the Matlab program:

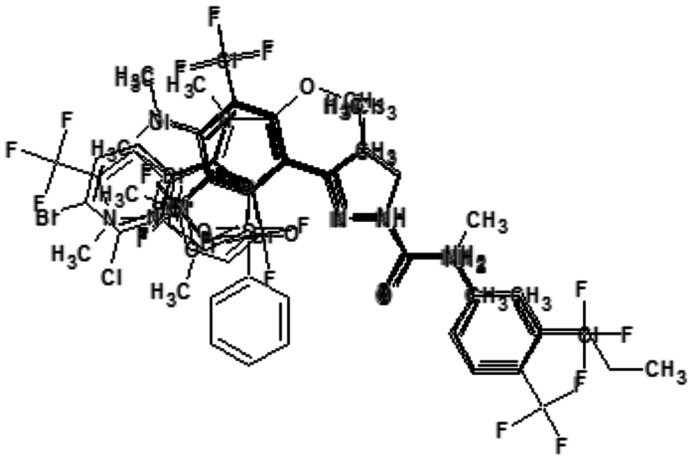
```
[name,MAP] = imread ('name.bmp','bmp');

name=double(name);

name=(name(:,:,1)+name(:,:,2)+name(:,:,3));
```

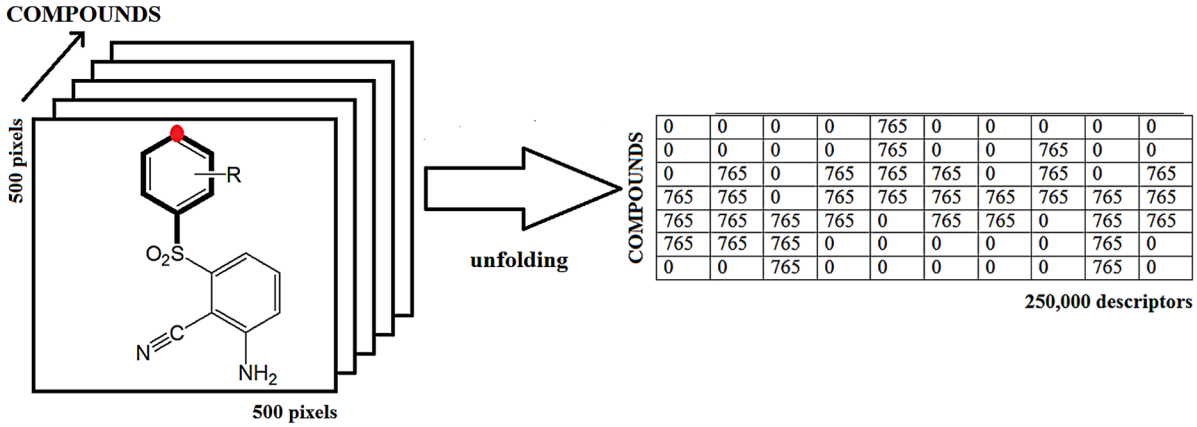
The n chemical structures of a data set, in which each one is described by a 2D image with $x \times y$ dimension, give a $n \times x \times y$ 3D array after superposition. This can be regressed against the **y** block (the bioactivities column vector) using a multilinear regression method, such as N-PLS. However, it is common to unfold the 3D array to an **X** matrix with $n \times (x \times y)$ dimension and then proceed with the regression using bilinear partial least squares (traditional PLS). This allows removing columns of the **X** matrix with zero variance, *i.e.* those columns containing pixel data corresponding to common substructures or common blank spaces; the time processing is quite improved, since thousands of descriptors are usually removed. An example of unfolded *binary* **X** matrix, *i.e.* a descriptors matrix containing 0 (black pixels forming each chemical structure drawing) and 765 (blank spaces in the drawing = white pixels) as molecular descriptors, is shown in Figure 3.

Figure 2. Superposition of a congeneric series of compounds illustrating the common substructure and the variable moieties of thiosemicarbazone derivatives with antitrypanosomal activity



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 3. Example of superposed workspaces with congeneric structures drawn and systematically aligned; the bold substructure is similar along with the whole congeneric series of arylsulfonyl compounds as anti HIV-1 compounds (Freitas, 2006). The alignment was performed by selecting the whole structure and adjusting a pixel in common among all of them (highlighted in red) in a given coordinate of the workspace. The resulting three-way array was unfolded to the binary **X** *matrix (containing 0 and 765 descriptors as the result of black and white pixels), which can be further regressed against the bioactivity values using PLS.*



0	0	0	0	765	0	0	0	0	0
0	0	0	0	765	0	0	765	0	0
0	765	0	765	765	765	0	765	0	765
765	765	0	765	765	765	765	765	765	765
765	765	765	765	0	765	765	0	765	765
765	765	765	0	0	0	0	0	765	0
0	0	765	0	0	0	0	0	765	0

Examples of successful MIA-QSAR modeling include the following series of compounds: a) (*S*)-*N*-[(1-ethyl-2-pyrrolidinyl)methyl]-6-methoxybenzamides as dopamine D_2 receptor subtype inhibitors, with $r^2 = 0.94$ and $q^2 = 0.58$ (Freitas, Brown, & Martins, 2005); b) 2-amino-6-arylsulfonylbenzonitriles and their thio and sulfinyl congeners as anti-HIV-1 compounds, with $r^2 = 0.81$, $q^2 = 0.71$ and $r^2_{test} = 0.82$ (Freitas, 2006); c) 3-phenyl-3-oxo-*N*-phenylpropanethiamides as antifungal compounds, with $r^2 = 0.96$ ad $q^2 = 0.62$ (Freitas & Rittner, 2008); d) sulfonylureas as herbicides, with $r^2 = 0.92$, $q^2 = 0.62$ and $r^2_{test} = 0.77$ (Bitencourt & Freitas, 2008).

QSAR models should be validated to probe for their reliability in predicting the bioactivities of potential drug candidates. While the leave-one-out cross-validation procedure is widely used for such a purpose, external validation has been considered the only way to achieve a reliable QSAR model (Golbraikh & Tropsha, 2002). Thus, the compound data set should be divided into training and test set compounds; the latter does not participate in the calibration step, but the activities of the compounds in this set should be predicted using the multivariate regression parameters. Usually, the statistical parameters required to consider a QSAR model as reliable are based on determination coefficients (r^2) ≥ 0.8 for calibration and ≥ 0.5 for external validation and leave-one-out cross-validation (also called q^2). However, additional criteria have been used to guarantee the reliability and sensitivity of QSAR models in various scenarios. These are mostly focused on the test set, whose bias concerned to scattering, slope and intercept of experimental *versus* predicted bioactivity values should be analyzed (Ojha et al., 2011; Chirico & Gramatica, 2012). Moreover, an **y**-randomization test guarantees that an eventual good calibration result is neither due to chance correlation nor the result of overfitting. Such a test is performed over the training set and a calibration is carried out using a shuffled **y**-block and the intact **X** matrix. A worse correlation in comparison with the real calibration model is expected if the descriptors really encode the biological properties. To evaluate the statistical difference between r^2 and $r^2_{y\text{-rand}}$, the ${}^c r_P^2$ should be calculated according to ${}^c r_P^2 = r \times (r^2 - r^2_{y\text{-rand}})^{1/2}$; values ≥ 0.5 are acceptable (Mitra, Saha, & Roy, 2010).

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

The existing MIA-QSAR models often respect these validation parameters, but more chemically representative MIA descriptors are expected to improve their quality and interpretability. For instance, the traditional MIA-QSAR method is based on thin wireframe-like structures; this may cause superposition deviation and imperfection during the 2D alignment. In addition, chemical symbols to indicate atoms (*e.g.* Br for bromine and C for carbon) are not chemically interpretable, despite being encoded in MIA-QSAR models. By the chance, it has been proved that MIA-QSAR really encodes biological data, since reliable models have been widely obtained, while images of letters of the alphabet do not correlate with the corresponding sequence number (Cormanich, Nunes, & Freitas, 2012), since this is arbitrary unlike MIA-QSAR. Even though, new dimensions were introduced in MIA-QSAR to increment quality and molecular encoding; the aug-MIA-QSAR method emerged.

Augmented MIA-QSAR

The new dimensions introduced in traditional MIA-QSAR to give the aug-MIA-QSAR method are based on varying colors and spheres with different radii to represent atoms. The molecules can be drawn with the aid of a program for molecular modeling and can be either optimized or not. Even though, aug-MIA-QSAR is essentially a 2D method because the images obtained from either 2D or 3D molecular structures are projections in the plane, a workspace with well-defined dimensions in which a common substructure along with the congeneric series is used for 2D alignment (superposition). The construction of the data matrix is practically the same as for the traditional MIA-QSAR, but the colors included in the aug-MIA-QSAR approach give rise to a variety of descriptors other than 0 (black pixels) and 765 (white pixels). Thus, while spheres with different sizes can be proportional to the van der Waals radii of atoms to represent steric effects, the variety of colors obtained from the sum of red, green and blue components can encode properties that differentiate atoms in the Periodic Table. For instance, the same series of compounds in Figure 2 was used to generate the structures to be used in the aug-MIA-QSAR modeling; superposition of these images is given in Figure 4 to have an idea of the data variance and drawing profile. The chemical structures were built using the GaussView program (2009) and the colors were kept as the program default.

One of the Periodic properties important in QSAR is the electronegativity, because it is related to bond polarity and, consequently, dipolar interactions useful to describe the substrate-enzyme interaction. Such information can be indicated by colors in aug-MIA-QSAR, since each atom color can be the result of the sum of red, green and blue contributions, in such a way that the resulting value (minimum of zero and maximum of 765) is proportional to the Pauling electronegativity scale. For example, the electronegativity values for fluorine, chlorine and carbon are 4.0, 3.2 and 2.6, respectively; the corresponding color composition to reflect this trend could be 600 (*e.g.* 255 red, 255 green, 100 blue) for fluorine, 480 (*e.g.* 200 red, 200 green, 80 blue) for chlorine, and 390 (*e.g.* 190 red, 100 green, 100 blue) for carbon. In this way, the 2D shape, atomic sizes and electronegativity are all encoded in the aug-MIA-QSAR method, which is illustrated in Figure 5. The comparison among all three image-based approaches (traditional MIA-QSAR, aug-MIA-QSAR with default atom colors, and aug-MIA-QSAR with atom colors proportional to electronegativity values) give insight about the limits of encoding of the traditional MIA-QSAR method, as well as the perspective for the chemical interpretation of images as molecular descriptors.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 4. Superposition of a congeneric series of compounds illustrating the common substructure and the variable moieties of thiosemicarbazone derivatives with antitrypanosomal activity, according to the aug-MIA-QSAR concept

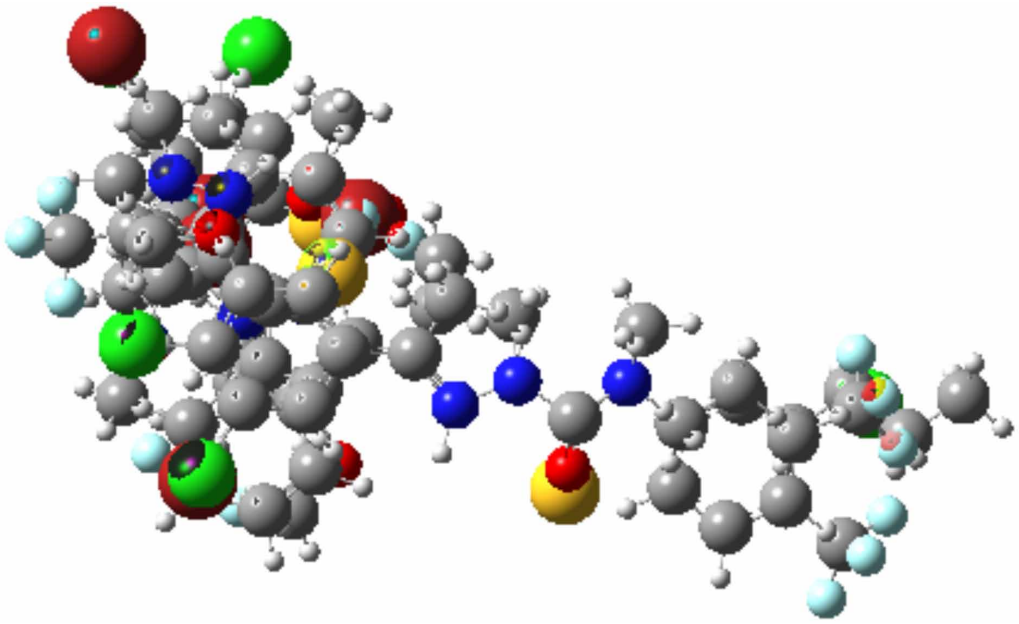
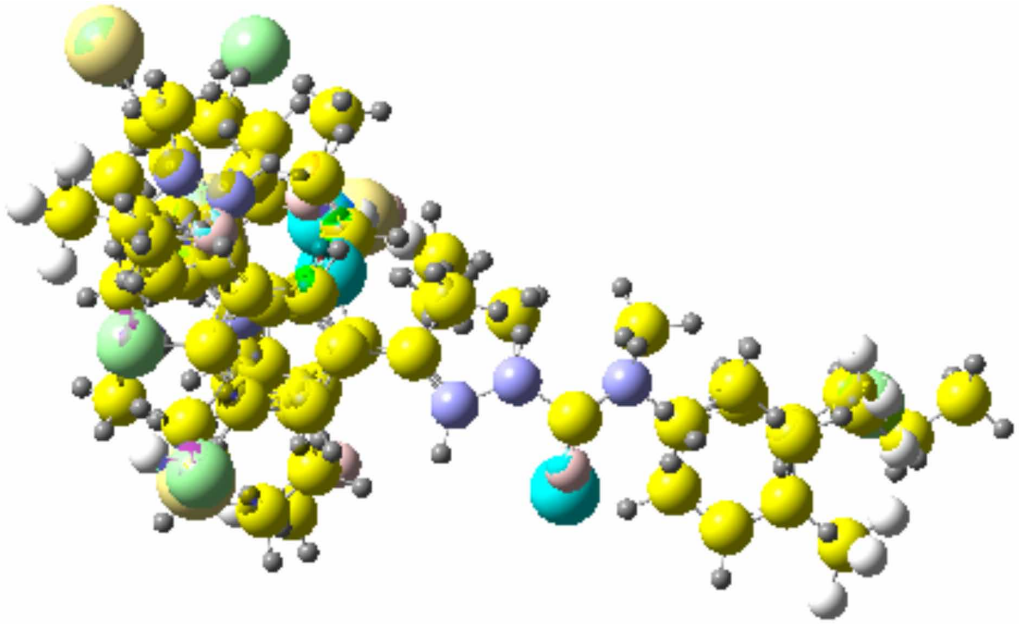


Figure 5. Superposition of a congeneric series of compounds illustrating the common substructure and the variable moieties of thiosemicarbazone derivatives with antitrypanosomal activity, according to the aug-MIA-QSAR concept using atom colors proportional to the Pauling electronegativity scale



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

CASE STUDY: TRYPANOSOMIASIS

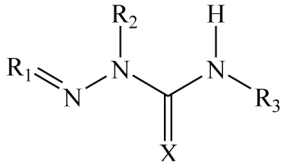
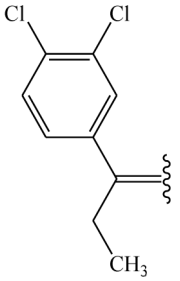
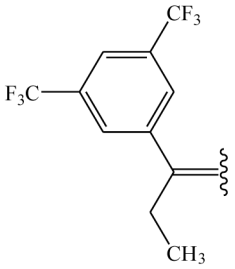
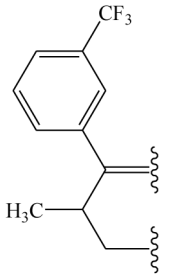
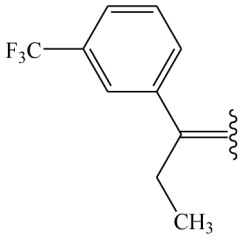
Despite the expected improvement of augmented MIA descriptors in comparison with traditional MIA-QSAR, there are only a few cases comparing the statistical advantages of using aug-MIA-QSAR over the traditional approach. Thus, this chapter reports a complete comparative analysis of calibration and validation parameters when using the aforementioned methods. Since QSAR methods are also useful to predict the bioactivity of non-existing drug candidates, the substructures of existing congeneric compounds were combined to give new drug-like compounds and their biological activities were estimated using the regression parameters obtained from all three MIA-based methods, in order to give insight about the numerical difference in pIC_{50} for these new compounds. The series of compounds comprises 53 thiosemicarbazones and semicarbazones acting as cruzain inhibitors extracted from the literature (Du et al., 2002; Lozano et al., 2012), which were divided into 43 training set compounds and 10 test set compounds (Table 1). The test set (*ca.* 20% of the whole series) was chosen in such a way that compounds with low, moderate and high activities were homogeneously distributed. The experimental values of pIC_{50} (IC_{50} is the half maximum inhibitory concentration, in mol L^{-1}) refer to the effectiveness against the *T. cruzi* cathepsin L-like protease (cruzain) using purified recombinant protein.

Because a large number of descriptors (generally thousands) is generated in MIA-based matrices, partial least squares regression is often used in the calibration step, thus requiring the choice of an optimum number of latent variables (PLS components) to proceed with the construction of the QSAR model. This can be achieved by analyzing the behavior of the root mean square errors of leave-one-out cross-validation (RMSECV) with increasing numbers of latent variables. An appropriate number of latent variables to be used in the model is that in which the RMSECV is minimized or does not vary significantly. The use of more latent variables than necessary can lead to overfitting, in which unnecessary/undesirable information is calibrated, giving good correlation coefficients in the calibration for the training set compounds, but not a reliable predictive model for external samples. The traditional MIA-QSAR method was applied for the series of compounds of Table 1 (the images corresponding to these compounds are superposed in Figure 2) and the optimum number of latent variables was 9, according to the plot of Figure 6.

The model was found to be somewhat overfitted, because the good correlation obtained in the calibration step ($r^2 = 0.967$, RMSEC = 0.162) is not significantly different from the results obtained for the PLS regression between the intact **X** matrix (the descriptors matrix) and the scrambled **y** block (the activities column vector), whose $r^2_{\text{y-random}}$ was 0.736 (mean of 10 repetitions). The statistical similarity between r^2 and $r^2_{\text{y-random}}$ can be evaluated by the aforementioned ${}^{c}r_{P}{}^{2}$ parameter, which was found to be inferior to 0.5 (0.473). In addition, despite of giving acceptable correlations in the leave-one-out cross-validation and external validation (Table 2), the low $r_{m}{}^{2}$ value obtained for the test set (0.382) indicates that the absolute predicted values of pIC_{50} does not match well the corresponding experimental data, *i.e.* the model experiences problems with the slope and/or intercept relative to the optimum line expected for the experimental *vs.* predicted pIC_{50}, which should pass through the origin. The correlation plots are given in Figure 7. In addition to experimental inaccuracy, statistical limitation, bad image alignment, etc., the source of such problem can be related to some lack of chemical information captured by the traditional MIA-QSAR approach. Thus, the QSAR model based on multivariate image analysis can be improved by introducing other dimensions to account for atomic sizes and other properties. This can be achieved carrying out aug-MIA-QSAR modeling over the set of compounds with antitrypanosomal activities.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

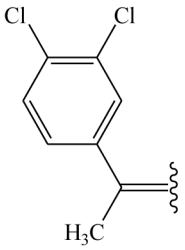
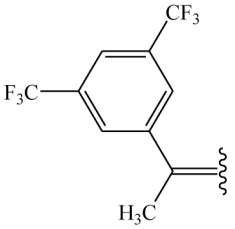
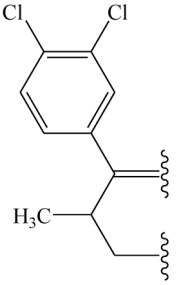
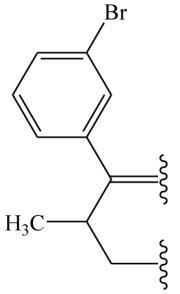
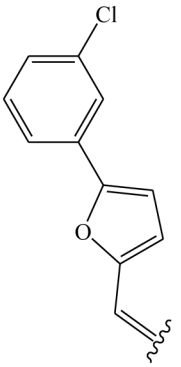
Table 1. Series of thiosemicarbazones and semicarbazones used in the QSAR modeling

Compound	R_1	R_2	R_3	X	Exp. Pic_{50}
					
1		H	H	S	7.70
2		H	H	S	7.70
3			H	S	7.40
4		H	H	S	7.30

continued on following page

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

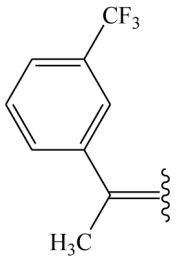
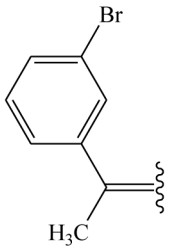
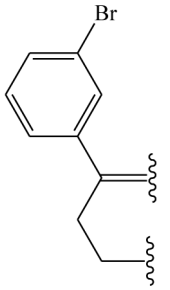
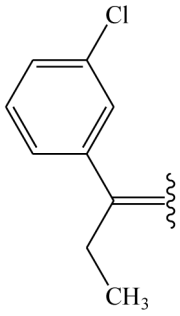
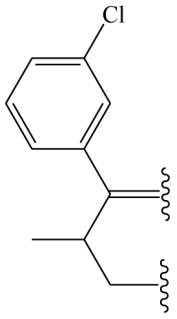
Table 1. Continued

Compound	R_1	R_2	R_3	X	Exp. Pic_{50}
5[a]		H	H	S	7.30
6		H	H	S	7.22
7			H	S	7.15
8			H	S	7.10
9		H	H	S	6.85

continued on following page

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

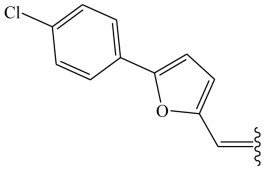
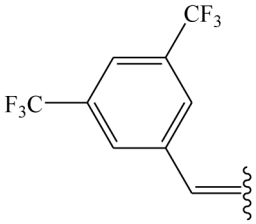
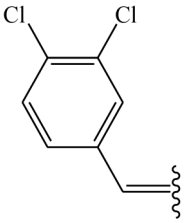
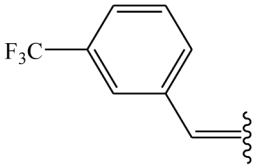
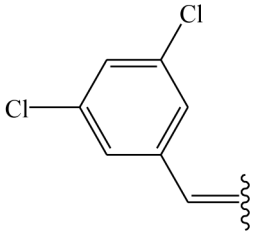
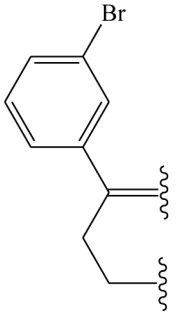
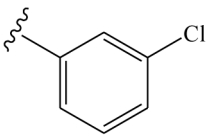
Table 1. Continued

Compound	R_1	R_2	R_3	X	Exp. Pic_{50}
10		H	H	S	6.77
11		H	H	S	6.70
12[a]			H	S	6.70
13		H	H	S	6.66
14			H	S	6.64

continued on following page

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

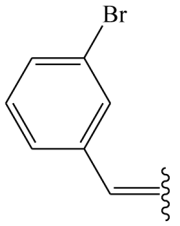
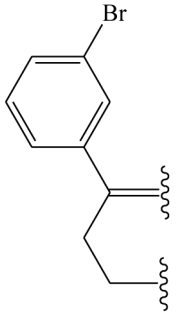
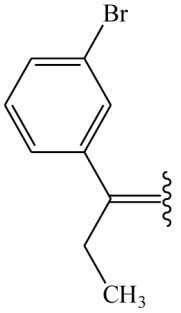
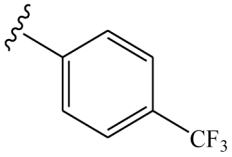
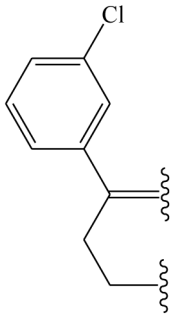
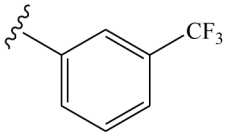
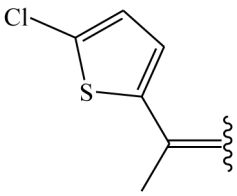
Table 1. Continued

Compound	R_1	R_2	R_3	X	Exp. Pic_{50}
15		H	H	S	6.52
16		H	H	S	6.49
17		H	H	S	6.34
18		H	H	S	6.32
19[a]		H	H	S	3.25
20				S	3.00

continued on following page

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

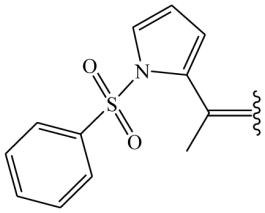
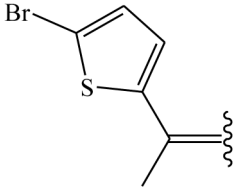
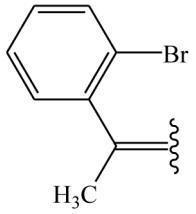
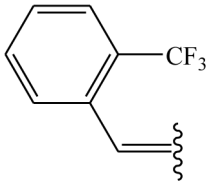
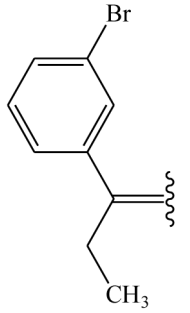
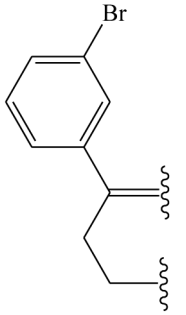
Table 1. Continued

Compound	R_1	R_2	R_3	X	Exp. Pic_{50}
21		H	H	S	5.89
22		H	$(CH_2)_5CH_3$	S	5.85
23[a]		H		S	5.80
24				S	5.77
25		H	H	S	5.72

continued on following page

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

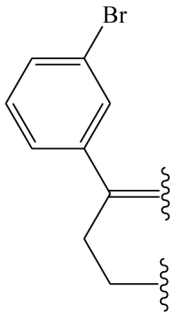
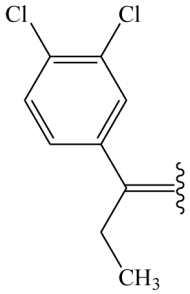
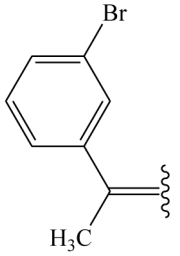
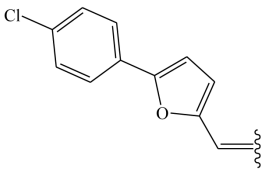
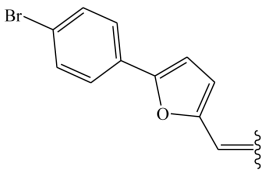
Table 1. Continued

Compound	R_1	R_2	R_3	X	Exp. Pic_{50}
26		H	H	S	5.55
27[a]		H	H	S	5.42
28		H	H	S	5.41
29		H	H	S	5.00
30		H	$(CH_3)_2$	S	5.00
31[a]			CH_3	S	5.00

continued on following page

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

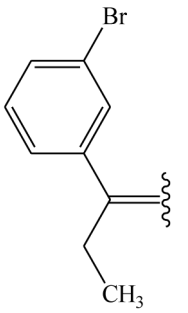
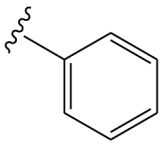
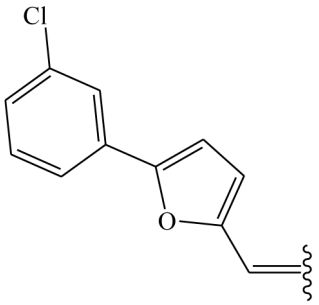
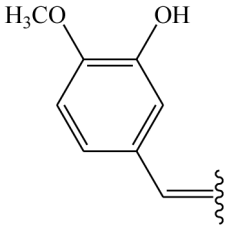
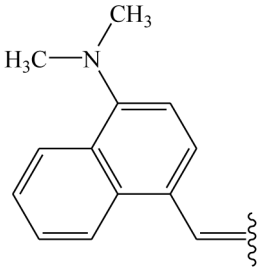
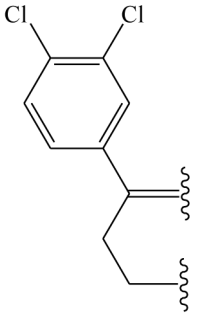
Table 1. Continued

Compound	R_1	R_2	R_3	X	Exp. Pic_{50}
32			$(CH_3)_2$	S	5.00
33		H	H	O	5.00
34		H	H	O	5.00
35		H	H	O	5.00
36[a]		H	H	O	5.00

continued on following page

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

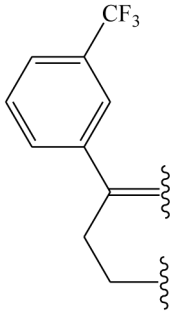
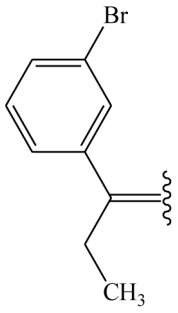
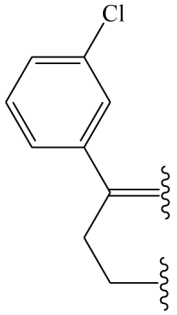
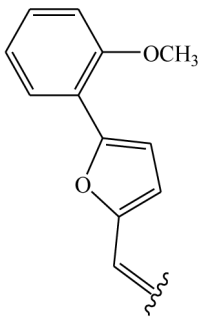
Table 1. Continued

Compound	R_1	R_2	R_3	X	Exp. Pic_{50}
37		H		S	5.00
38		H	H	O	5.00
39		H	H	S	5.00
40		H	H	S	5.00
41			H	S	7.22

continued on following page

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

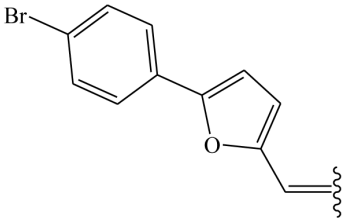
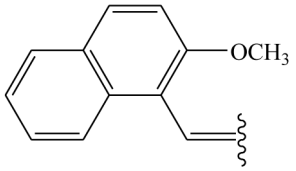
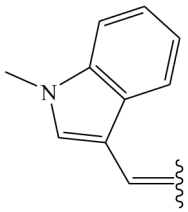
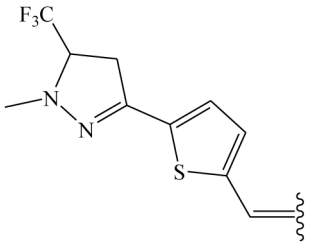
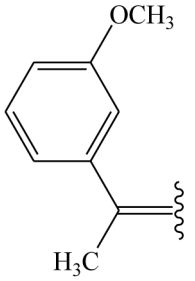
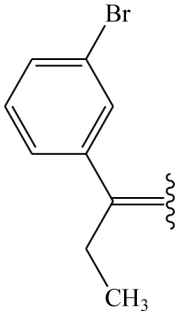
Table 1. Continued

Compound	R_1	R_2	R_3	X	Exp. Pic_{50}
42[a]			H	S	7.10
43		H	H	S	7.00
44			H	S	6.57
45		H	H	S	6.55

continued on following page

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

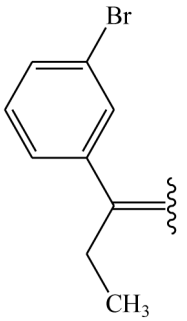
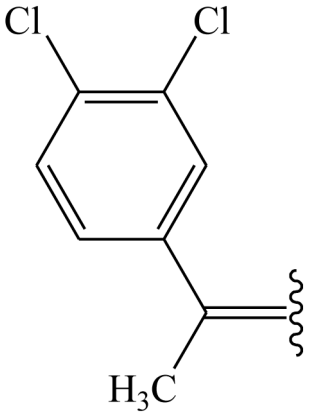
Table 1. Continued

Compound	R_1	R_2	R_3	X	Exp. Pic_{50}
46[a]		H	H	S	6.55
47		H	H	S	6.25
48		H	H	S	5.55
49		H	H	S	5.15
50[a]		H	H	S	5.00
51		H	CH_3	S	5.00

continued on following page

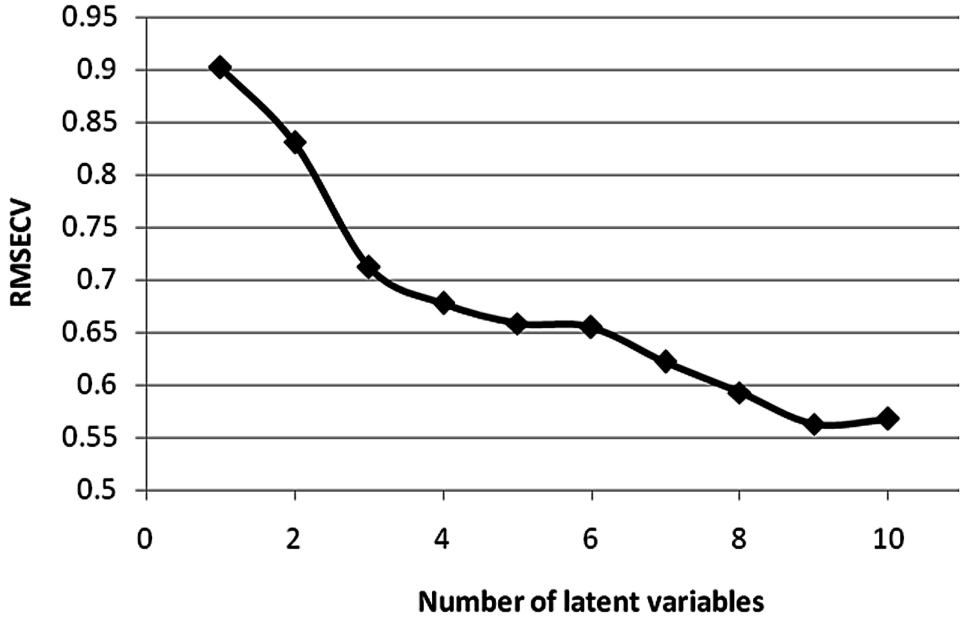
EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 1. Continued

Compound	R_1	R_2	R_3	X	Exp. Pic_{50}
52		H	CH_2CH_3	S	5.00
53		H	H	O	5.00

[a]Test set compounds.

Figure 6. Plot of number of latent variables vs. RMSECV for the traditional MIA-QSAR model



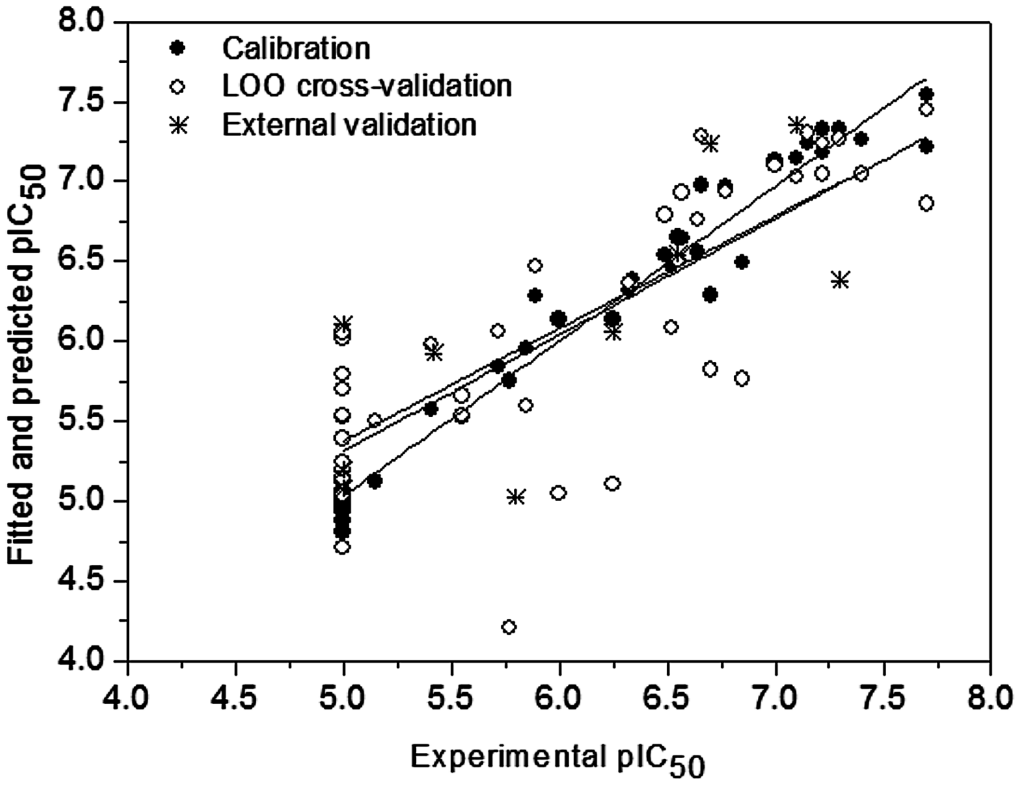
EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 2. Statistical results for the QSAR models obtained for the series of antitrypanosomal compounds

Parameter	Traditional MIA-QSAR	Aug-MIA-QSAR	Aug-MIA-QSAR$_{color}$
Latent variables	9	9	9
r^2	0.967	0.966	0.966
RMSEC	0.162	0.166	0.164
q^2	0.624	0.763	0.756
RMSECV	0.563	0.443	0.452
r^2_{test}	0.569	0.654	0.714
RMSEP	0.579	0.512	0.480
r_m^2	0.382	0.481	0.594
$r^2_{y\text{-}random}$ [b]	0.736	0.702	0.696
[c]r_P^2	0.473	0.504	0.511

[a]Aug-MIA-QSAR$_{color}$ refers to the model in which colors are proportional to the electronegativities. [b] Mean of 10 repetitions.

Figure 7. Plot of experimental vs. fitted and predicted pIC_{50} for the series of thiosemicarbazones and semicarbazones, according to the traditional MIA-QSAR method

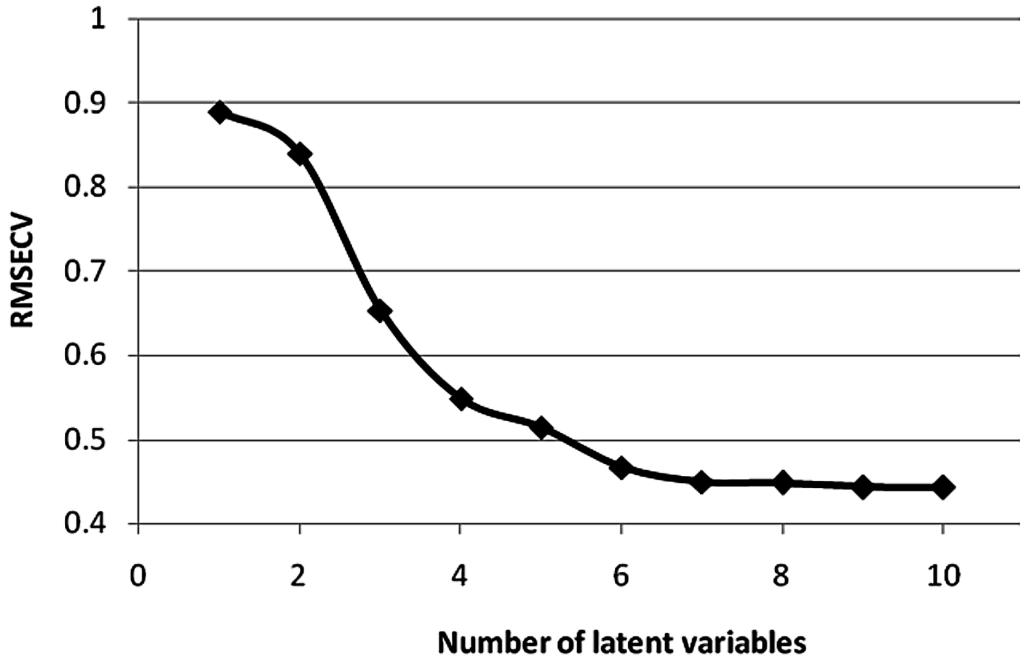


EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

The aug-MIA-QSAR method was applied to model the bioactivities of the same class of compounds evaluated using traditional MIA-QSAR, in order to get insight about the statistical advantages of using this augmented methodology. The chemical structures were built using the GaussView program, in which default colors for each atom are those represented in Figure 4 for the superposed images (gray for carbon, white for hydrogen, light blue for fluorine and so on). These colors are arbitrary and do not have any correspondence with atomic properties, *i.e.* the RGB additive color model gives 426 for carbon (resulting in the gray color as a composition of 142 red, 142 green and 142 blue), 612 for hydrogen (resulting in the white color as a composition of 204 red, 204 green and 204 blue), 688 for fluorine (resulting in the light blue color as a composition of 178 red, 255 green and 255 blue), and so on. Thus, different colors for different atoms serve to distinguish the spheres with different sizes in the images, but they do not have chemical meaning. On the other hand, spheres with sizes proportional to the van der Waals radii describe the atomic volume and corresponding properties, such as steric effect. The effect of these changes over the traditional MIA-QSAR model was investigated.

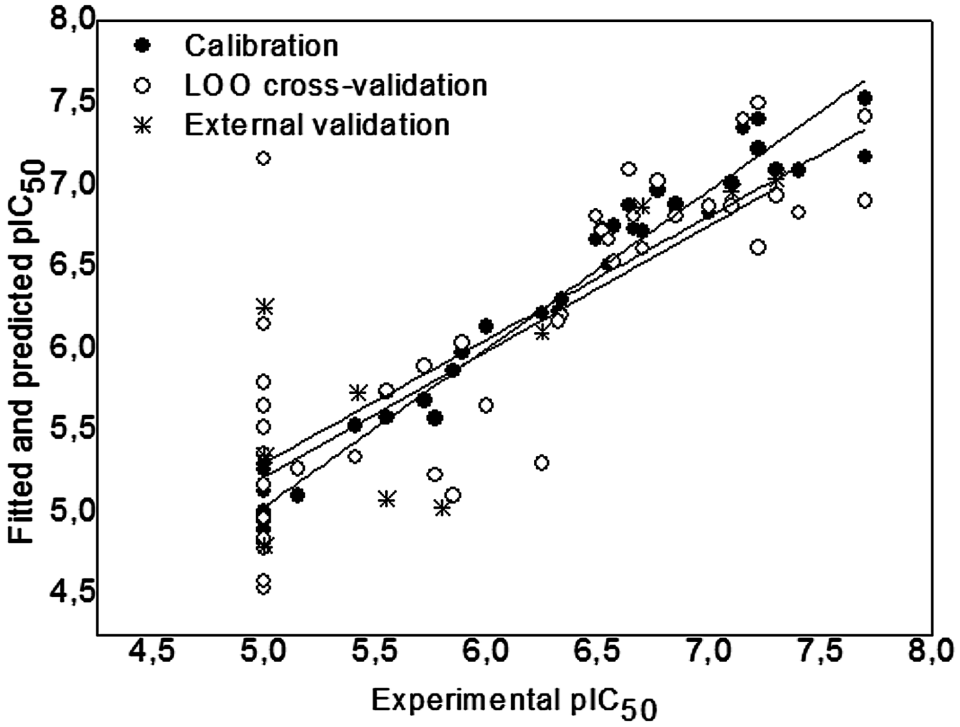
The same procedure described to build the traditional MIA-QSAR model was applied to construct the aug-MIA-QSAR model. Again, 9 PLS components were chosen to be used in the modeling; the RMSECV appears to stabilize from 7 latent variables (Figure 8), but 9 were used to reproduce the conditions applied in the traditional MIA-QSAR model. The calibration gave a high correlation between experimental and fitted pIC_{50} values, while the results did not show to be overfitted, since the ${}^{c}r_{P}^{2}$ value was superior to 0.5 (Table 2). In addition, the leave-one-out cross-validation was improved, with q^2 of 0.763, which is quite satisfactory. However, external validation has been claimed as the only way to establish a reliable QSAR model (Golbraikh & Tropsha, 2002) and, despite the good correlation between experimental and predicted pIC_{50} values for this set of compounds (Figure 9), it did not pass the r_m^2 test, whose value was marginally inferior to 0.5. Thus, the aug-MIA-QSAR model also requires improvement to certify that predictions for new drug candidates are really reliable.

Figure 8. Plot of number of latent variables vs. RMSECV for the aug-MIA-QSAR model



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 9. Plot of experimental vs. fitted and predicted pIC_{50} for the series of thiosemicarbazones and semicarbazones, according to the aug-MIA-QSAR method (default colors)

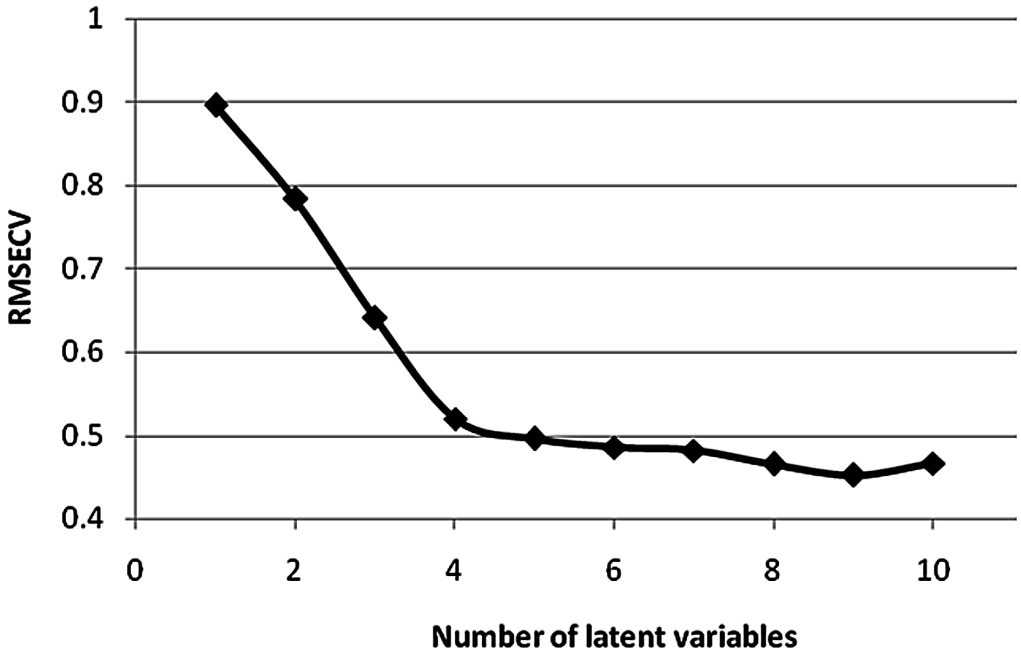


A new dimension can be introduced in aug-MIA-QSAR if colors in the chemical images are used to encode atomic properties. In addition to steric effects, other important interactions ruling enzyme-substrate interactions, responsible for the activity of drug molecules, are the dipolar interactions. Only atoms involved in polar bonds are capable of interacting through dipolar interactions. A polar bond is formed by atoms with different electronegativities and, therefore, the Pauling electronegativity can be used as a descriptor in the so called aug-MIA-$QSAR_{color}$ method. According to this approach, different atoms are colored according to the respective electronegativity values. Because colors can be numerically managed using the RGB additive color model, the atom colors can be proportional to the corresponding electronegativity values, *e.g.* in Figure 5, where the fluorine with electronegativity 4.0 is approximately white (760 as a sum of 255 red, 255 green and 250 blue), the carbon with electronegativity 2.5 is approximately yellow (475 as a sum of 255 red, 225 green and 0 blue), the hydrogen with electronegativity 2.1 is approximately gray (399 as a sum of 133 red, 133 green and 133 blue), and so on.

Similarly to the previous aug-MIA-QSAR, the aug-MIA-$QSAR_{color}$ could be more parsimonious than the traditional MIA-QSAR model, since the RMSECV decayed on going from 1 to 5 latent variables, from which it is stabilized (Figure 10). However, 9 PLS components were used in the aug-MIA-$QSAR_{color}$ modeling for comparison with the other models. Here, the calibration results continued statistically significant, with low RMSEC, high r^2 and not overfitted ($^c r_P^2 > 0.5$). Nevertheless, the main advantages of this new model is related to its validation, because the leave-one-one cross-validation results suggest a high predictability ($q^2 = 0.756$), which is confirmed by the external validation, in which r^2_{test} was improved to 0.714, while the absolute experimental and predicted pIC_{50} values matches very well to each

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 10. Plot of number of latent variables vs. RMSECV for the aug-MIA-QSAR model (colors proportional to the atom electronegativities)



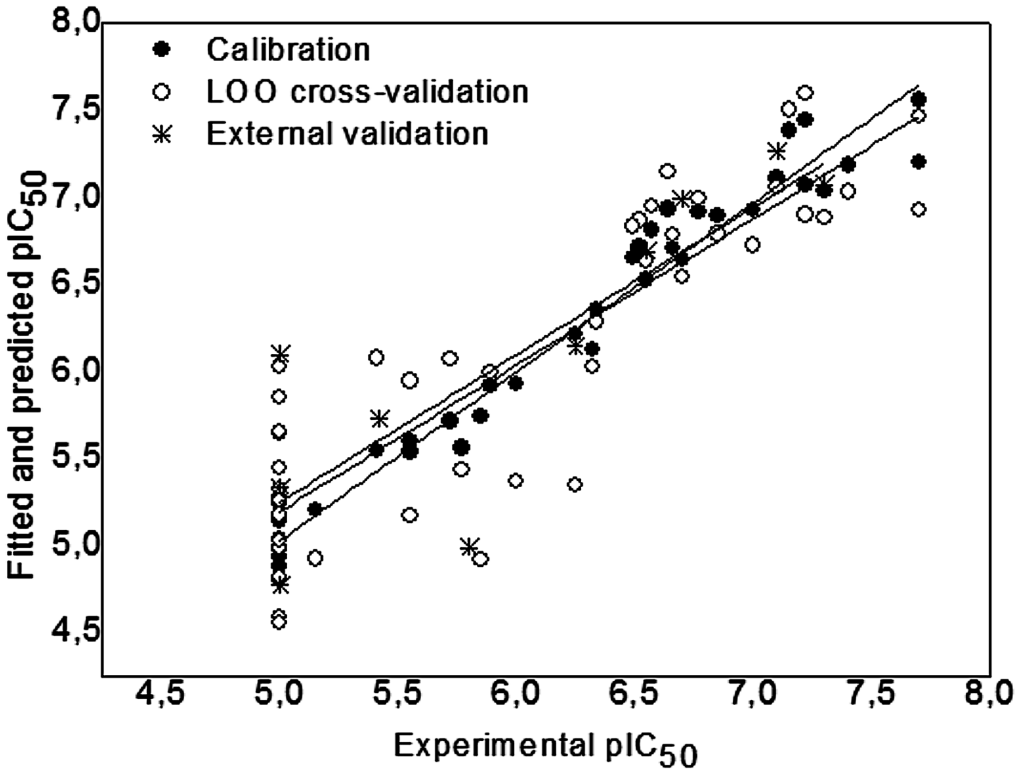
other ($r_m^2 > 0.5$). The correlations are illustrated in Figure 11. Now, the aug-MIA-QSAR$_{color}$ model for the series of thiosemicarbazones and semicarbazones as antitrypanosomal compounds satisfies the most important validation parameters required to attest the reliability of a QSAR model, suggesting that the chemical properties correlating with biological activities are suitably encoded by the aug-MIA$_{color}$ descriptors and that dipolar interactions play a key role for the action mechanism of this series of compounds. Consequently, the aug-MIA-QSAR$_{color}$ model is ready to estimate the bioactivities of new drug candidates.

A QSAR analysis based on physicochemical descriptors for the same series of compounds has been performed, but using a different set of test compounds. The best PLS model using 8 selected descriptors gave $r^2 = 0.85$, $q^2 = 0.78$ and $r^2_{test} = 0.55$ (Lozano et al., 2012). Thus, the MIA-based models are consistent with the classical QSAR using physicochemical descriptors, despite a deeper comparison cannot be done because the lack of additional validation in the later.

A useful strategy to obtain new compounds containing a given scaffold is to combine the molecular substructures of highly active compounds in a congeneric series. This can be achieved using the data set of Table 1. This procedure can result in highly active substances with improved properties, such as lower toxicity and high activity toward strains that are resistant against the available chemotherapy. For example, compounds **1** and **2** of Table 1 are the most active ones in that series, but they differ from each other by the substituents at the benzene ring in $\mathbf{R_1}$ (3,4-Cl_2 in **1** and 3,5-$(CF_3)_2$ in **2**). Thus, a new drug-like compound can be designed with $\mathbf{R_1}$ containing the benzene ring with 3-Cl,5-CF_3 or 4-Cl,5-CF_3 or 3-CF_3,4-Cl or 3,4-Cl_2,5-CF_3 or 3,5-$(CF_3)_2$,4-Cl; all these possibilities have the Cl and CF_3 groups calibrated in that ring positions. Accordingly, these compounds (**A-E**, Figure 12) were proposed and their bioactivities were subsequently estimated using the regression parameters of the MIA-QSAR and aug-MIA-QSAR models.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 11. Plot of experimental vs. fitted and predicted pIC_{50} for the series of thiosemicarbazones and semicarbazones, according to the aug-MIA-$QSAR_{color}$ method (colors proportional to the atom electronegativities)



According to the traditional MIA-QSAR model, compounds **A** and **C** were the most active purposes (Table 3), but not more active than the experimentally available compounds **1** and **2**. However, the most predictive aug-MIA-QSAR models indicated that **C** and **D** are more active than **1** and **2**. Consequently, these proposed compounds can drive the synthesis of new antitrypanosomal drugs. In addition, these outcomes allow the chemical interpretation about structural requirements to achieve successful purposes: thiosemicarbazones containing a tri-substituted phenyl ring in the $\mathbf{R}_1$ group are promising, while chlorine at position 3 is preferred rather than the trifluoromethyl group.

These results can be further validated by comparing the predicted pIC_{50} values with the docking scores obtained from receptor-based methods. Molecular docking predicts the preferred orientation of a molecule (substrate) to another (enzyme) when bound to each other to form a stable complex. This binding affinity is usually described in terms of scoring functions, which are expected to be correlated with the bioactivity values. Thus, structure- and receptor-based approaches are complementary. A comparison between docking scores and the bioactivity data obtained through several MIA-QSAR models (Antunes et al., 2008; Goodarzi et al., 2010; Deeb et al., 2012; Guimarães et al., 2014) has shown that MIA descriptors indeed carry biochemical information. Such a comparison has not performed hitherto using aug-MIA descriptors, but the expected response is optimist, since the predictive performance of aug-MIA-QSAR is better than in traditional MIA-QSAR, at least for the case study of this chapter.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 12. Proposed thiosemicarbazones obtained from the combination of the substructures of compounds 1 and 2

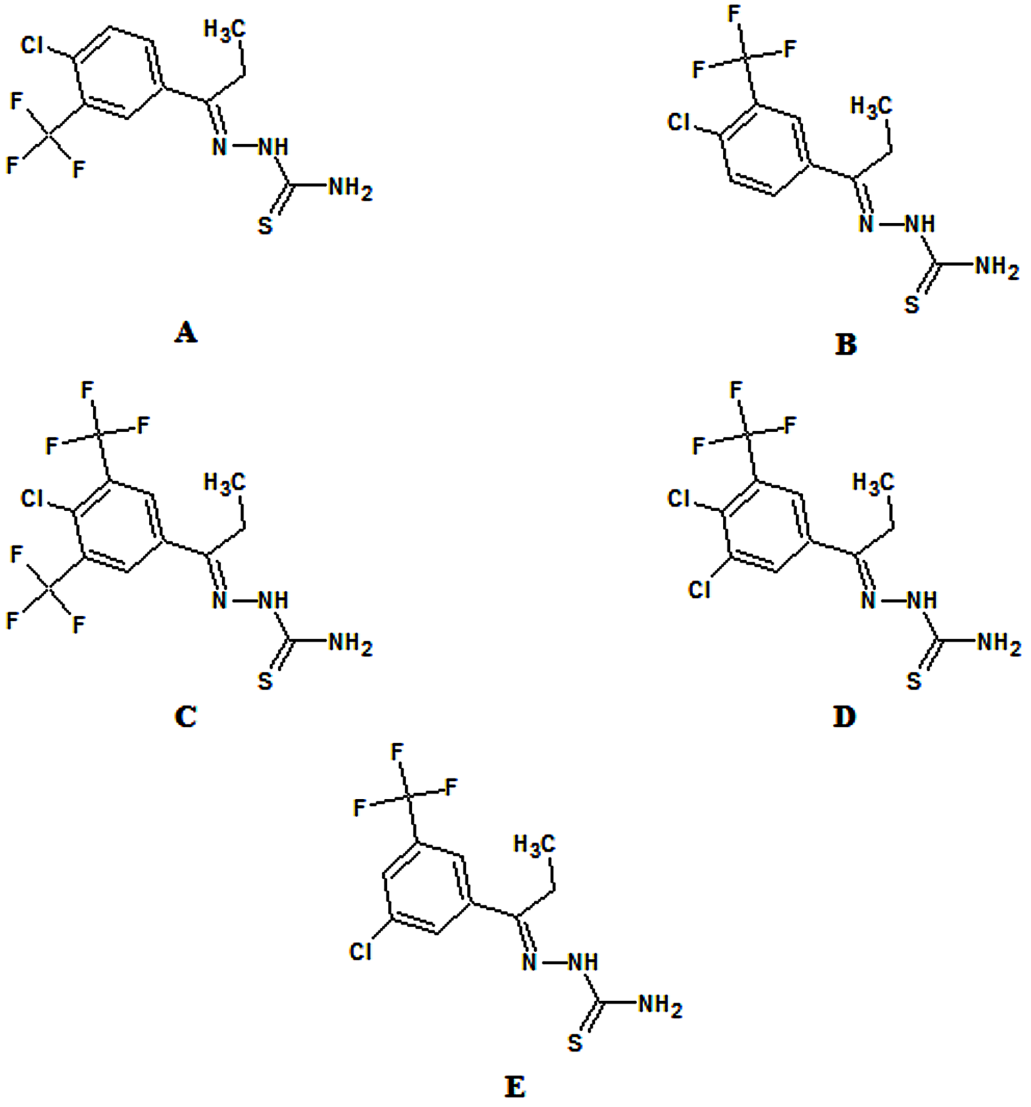


Table 3. Predicted pIC_{50} for the proposed thiosemicarbazones **A-E**

Compound	Traditional MIA-QSAR	Aug-MIA-QSAR	Aug-MIA-QSAR$_{color}$
A	7.37	7.55	7.48
B	6.73	7.38	7.44
C	7.60	7.98	8.01
D	7.07	7.76	7.76
E	6.95	7.30	7.31

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Additional drug-like profile can be estimated by numerous programs available for this purpose, which have also complemented many pharmacodynamic data obtained by MIA-QSAR (Silva et al., 2012; Guimarães et al., 2014). The estimation of pharmacokinetic data, namely ADMET (absorption, distribution, metabolism, excretion and toxicity) can be directly and indirectly calculated. The most widely known indirect method to obtain some information about drug-likeness of a given drug candidate is to use the Lipinski's rule of five (Lipinski et al., 1997). According to this rule, most drug-like molecules have log$P \leq 5$, molecular weight ≤ 500, number of hydrogen bond acceptors ≤ 10, and number of hydrogen bond donors ≤ 5. Molecules violating more than one of these rules might have problems with bioavailability. The rule is called "rule of 5", because the border values are 5, 500, 2×5, and 5. For example, while hydrophobicity (expressed in terms of log$P$) affects drug absorption, bioavailability, hydrophobic drug-receptor interactions, metabolism of molecules, as well as their toxicity, the molecular volume determines transport characteristics of molecules. Obviously, this is an approach, since many current therapies are based on drug molecules violating more than one of these rules, such as Ritonavir, which is used for the treatment of AIDS, but violates 3 rules according to calculations using the Molinspiration program (www.molinspiration.com), i.e. logP > 5, molecular weight of 721 g mol^{-1} and 11 hydrogen bond acceptors. Other programs estimate directly some pharmacokinetic parameters, such as human intestinal absorption and the blood-brain barrier penetration. These calculations are mostly based on models comprising big libraries of compounds having such parameters available.

In fact, the combination of a QSAR technique (to predict the bioactivity of a designed compound) with docking studies (to understand the substrate-enzyme interaction mechanism) and ADMET evaluation, is a powerful way to achieve a successful drug candidate in a relatively low cost manner and more quickly than performing experimental compound screening and tests. Thus, multivariate image analysis methods can play an important role on this course.

FUTURE RESEARCH DIRECTIONS

The MIA-based QSAR methods are relatively new and, consequently, many performance tests and improvements are still to be performed, while the scientific community and the whole society demand more and better medicines to treat increasingly numerous and damaging diseases. This fact makes necessary strategies in medicinal chemistry to provide suitable chemotherapies in a fast, reliable and safe manner. In order to satisfy these requirements, new computational methods are always welcome to be developed, as are mathematical tools to capture the maximum information possible from molecular descriptors and additional validation techniques to guarantee that QSAR models are really predictive and accurate. Thus, future research directions related to MIA-QSAR are concerned to improve the biochemical interpretability of MIA descriptors, to mathematically manage descriptors in order to capture the MIA information not explored using conventional tools and to apply more precautionary approaches for validation purposes.

There is a well-known lack of interpretability in traditional MIA-QSAR. Partially, this can be attributed to the binary-like descriptors and the big three-way data array obtained after images superposition. This means that the three-way array should be unfolded to a two-way array (a matrix) in order to be analyzed by bilinear multivariate methods, such as PLS (for regression) and PCA (for pattern recognition). The unfolding makes difficult to identify which descriptor corresponds to which pixel coordinate in the 2D space forming each image. In addition, in order to make the data management more practicable, in terms of computational cost, columns with zero variance are usually removed; this procedure makes calculation

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

faster, but the information about which number (765 for white pixels and 0 for black pixels) in the data matrix corresponds to which pixel location in the image gets completely lost. In addition, back-folding becomes impossible. This drawback can be at least partially overcome using aug-MIA-QSAR, because the descriptor number corresponding to a given pixel in a chemical image can be identified (since each atom has a different color) and its contribution for the model can be investigated using loading analysis. In aug-MIA-QSAR$_{color}$, this information is expected to be still more chemically meaningful. This is inspired in the 3D contour maps obtained from CoMFA analysis, where different colors are given to surfaces indicating positive, negative or neutral effect of certain descriptors (e.g. steric and electrostatic) on the bioactivity values. Color is also used to code the direction and magnitude of these differential interactions and, in CoMFA, the numerical data used to construct these "coefficient contour" maps, the QSAR coefficients and the data table, are available on request from the authors (Cramer, Patterson, & Bunce, 1988).

Even the more chemically meaningful aug-MIA-QSAR$_{color}$ method (compared to traditional MIA-QSAR and aug-MIA-QSAR) may not offer descriptors linearly correlated with the biactivity values, or then some information may not have any significance at all; in the worst-case scenario, biased descriptors (e.g. from spurious data, bad alignment, etc.) could confuse the model. Thus, variable selection and different regression methods can be used to overcome these setbacks. Genetic algorithm is one of the most widely applied methods for feature selection. It has widespread use for dimension reduction and have been reported in the optimization of a number of different and traditionally difficult problems, including image processing, design of complex networks (e.g., computers and integrated circuits), classifications, parameters for neural nets, job scheduling, robotics, and parameter fitting (Davis, 1991). Genetic algorithm (GA) is a search that mimics the process of natural selection and it is based on the Darwin's evolutionary theory; a population of candidate solutions (compounds) to an optimization problem is evolved toward better solutions. Each candidate solution has a set of properties (descriptors) which can be mutated and altered. The evolution usually starts from a population of randomly generated compounds, and is an iterative process, with the population in each iteration called a *generation*. In each generation, the fitness of every compound in the population is evaluated; the fitness is usually the value of the objective function in the optimization problem being solved. The more fit individuals are stochastically selected from the current population, and each individual's *genome* is modified (recombined and possibly randomly mutated) to form a new generation. The new generation of candidate solutions is then used in the next iteration of the algorithm. Commonly, the algorithm terminates when either a maximum number of generations has been produced, or a satisfactory fitness level has been reached for the population (Whitley, 1994). A variety of other feature selection methods is available, from those based on ordered predictors selection (OPS), which consists of obtaining an informative value, then decreasing the ordering of variables by this value, and lastly investigating the ordered variables (Teófilo, Martins, & Ferreira, 2009), to those inspired in ant colony optimization (Jensen, 2006). The use of such methods can be particularly useful if applied to aug-MIA descriptors, since the selected, colored descriptors could be identified and then analyzed on the basis of their contribution for the activity values.

Bilinear partial least squares (PLS) is the main regression method used in multivariate QSAR. In MIA-QSAR, unfolding is done so that pixels become a single row and, thus, an image that is originally I by J pixels for K compounds is reshaped to form a two-way array that is I×J by K. An **X** matrix is then built, where each row contains the variables (the pixels) describing each molecule, and is subsequently decomposed into a score vector **s** and a weight vector **w**. The score vector is determined to have the property of maximum covariance with the dependent variable **y**. The score vectors then replace the original

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

variables as regressors. However, the three-way array can be directly regressed against the **y** column vector using multi-way regression methods, such as multilinear partial least squares (N-PLS) (Bro, 1996) and parallel factor (PARAFAC) analysis (Bro, 1997). Actually, these methods have not shown much improved results if compared to PLS when applied to traditional MIA descriptors (Freitas et al., 2008). On the other hand, regression methods that account for nonlinearities, particularly artificial neural networks (ANN) and support-vector machines (SVM), have really enhanced the predictive performance of physicochemical-based QSAR models when compared to multiple linear regression (Goodarzi, Freitas, & Jensen, 2009). Least-squares support-vector machines (LS-SVM) combined with feature selection methods have been applied only in a few cases to MIA descriptors (Cormanich, Goodarzi, & Freitas, 2009); thus, research on this field is a promising perspective to achieve better (aug)MIA-based models where the relationship between structure and activity is not completely linear.

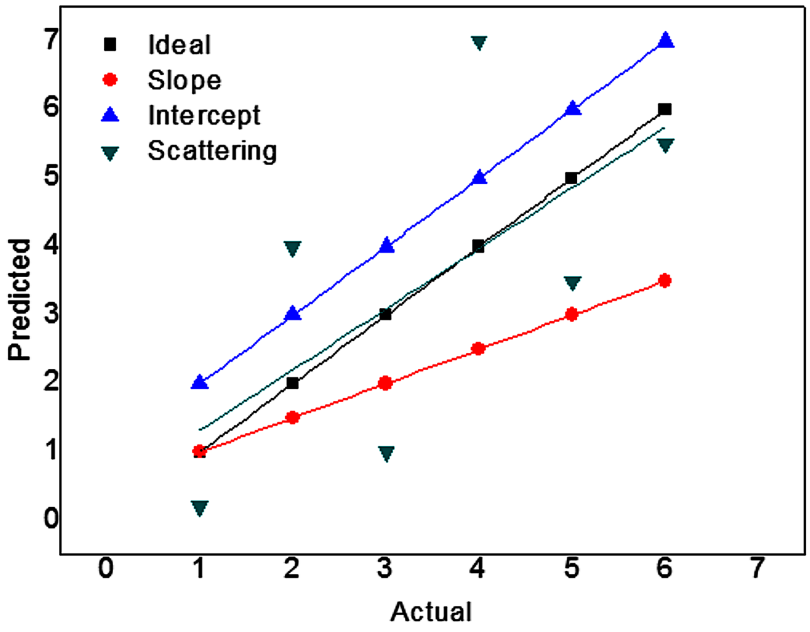
A QSAR model should be reliable enough to allow the prediction of external samples in order to lead the synthesis and development of new drug candidates. Such a confidence can be assessed using the maximum number possible of validation protocols. There are several examples of traditional MIA-QSAR models in the literature submitted to validation tests, such as leave-one-out cross-validation, external validation, **y**-randomization tests, etc. The results have shown to be satisfactory. The leave-one-out cross-validation (LOOCV) is the most widely used procedure for validation purposes; LOOCV involves using a single observation from the original set of compounds as the validation data, and the remaining observations as the training data. This is repeated such that each observation in the set of compounds is used once as the validation data. Alternatively, more than a single observation can be used in each observation, in the so called leave-many-out cross-validation. The main result is the determination coefficient q^2 obtained from the correlation between experimental *versus* predicted data. However, the use of LOOCV alone for validation purposes has been criticized (Golbraikh & Tropsha, 2002), while external validation has been claimed as the only way to establish a reliable QSAR model.

An external validation test is performed to a series of compounds different from those pertaining to the training set, which is used for calibration purposes. The regression parameters obtained from the calibration process are used to predict the dependent variables of the external set of compounds. The main result is the determination coefficient r_{test}^2 obtained from the correlation between experimental *versus* predicted data. In general, r_{test}^2 above 0.5 are considered acceptable, but it is not a sufficient condition to attest the reliability of a QSAR model. Therefore, additional statistical parameters applied over the external sample of compounds have been developed to prevent misinterpretation of an eventual high determination coefficient of external validation. There are different sources of problems in the external validation that can lead to equivocal consideration about the suitability of a QSAR model, namely data scattering and deviation of the intercept and slope of the line obtained from the linear regression of actual *versus* predicted data, in comparison with an ideal scenario (Figure 13).

Highly scattered data could result in a regression line almost coincident with the one obtained from an ideal scenario, but the problem can be captured by the low r_{test}^2 value and high root mean square errors (RMSE). However, there are cases in which r_{test}^2 is high if the regression line is not forced to pass through the origin, but the QSAR model is not really predictive. For instance, external validation data in which the predicted values present a systematic error compared with the actual ones will have problem with the line intercept. In another case, a correlation giving high r_{test}^2 value could be obtained from a relationship in which the regression line passes through the origin, but this will have problem with the slope of the curve if the absolute values of actual and predicted data are not closely related. These and other problems can be analyzed using a variety of statistical parameters, like the modified r^2

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 13. Examples of possible problems in external validation compared with an ideal scenario



(r^2_m), Δr^2_m and corrected-penalized r^2 ($^c r^2_P$) of Equations 1-3 (Roy et al., 2009; Roy et al., 2012; Roy et al., 2013; Ojha et al., 2011; Mitra, Saha, & Roy, 2010). While $r^2_m > 0.5$ guarantees that not only a good correlation between the experimental and predicted values is obtained for the test set, but also that the absolute experimental and predicted pEC_{50} values are congruent, the $^c r^2_P$ parameter gives insight about the statistical difference between r^2 and r^2_{rand} (values above 0.5 are considered acceptable).

$$r_m^{\,2} = r^2 \left[1 - (r^2 - r_0^{\,2})^{1/2}\right] \quad (1)$$

wherein r^2 and $r_0^{\,2}$ correspond to the squared correlation coefficient values between measured and predicted values for the test set with and without intercept, respectively. Change of the axes gives the value of r'^2_m, which is used in Equation 2 to obtain Δr^2_m ($\Delta r^2_m < 0.2$ is acceptable).

$$\Delta r^2_m = |r^2_m - r'^2_m| \quad (2)$$

$$^c r^2_P = r\,(r^2 - r^2_{rand})^{1/2} \quad (3)$$

k and k' indicate the slopes of the observed *versus* predicted data (Equations 4 and 5) and are useful for further validation. Ideally, both k and k' should be ≥ 0.88 and ≤ 1.15.

$$k = \Sigma(y_{obs} \times y_{pred})/\Sigma(y_{pred})^2 \quad (4)$$

$$k' = \Sigma(y_{obs} \times y_{pred})/\Sigma(y_{obs})^2 \quad (5)$$

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

More precautionary criteria have been recently established by Chirico and Gramatica (2012), based on statistical parameters obtained from the calibration, cross-validation and external validation sets. Overall, the authors suggest that, in addition to the most commonly used validation criteria to different kinds of bias in data, the concordance correlation coefficient (CCC) is an important parameter to verify how small the differences are between experimental data and external data set predictions, independently of their range. Because aug-MIA-QSAR is a relatively new method, further research to probe for its adequacy to these trends in validation is a perspective in this field.

CONCLUSION

Multivariate image analysis applied in quantitative structure-activity relationship studies emerged as a necessity to design new drug candidates in a reliable and easy way. This can be demonstrated by a variety of examples in the literature using traditional MIA-QSAR in the prediction of bioactivities of proposed compounds. This chapter described the procedure used to build a MIA-QSAR, as well as its evolution from 2005 to the current augmented versions (aug-MIA-QSAR), which encode chemical information better than the traditional MIA-QSAR. The molecular representation was improved by introducing spheres with different sizes and colors corresponding to the atoms in molecules. While the sphere sizes are expected to account for biological effects depending on the atomic (or group of atoms) volume, like steric effects, the colors could encode electrostatic interactions, especially if sphere colors are numerically proportional (according to the RGB color model) to the electronegativity Pauling scale. This improvement was checked for a series of thiosemicarbazones and semicarbazones as antitrypanosomal agents; the so called aug-MIA-QSAR$_{color}$ model was found to be more predictive and reliable than other MIA-based models. However, despite promising, such method is quite recent and then lacks additional tests (development of models for several classes of compounds and comparisons), improvements concerned to data analysis and interpretability, and application of state-of-the-art validation protocols. Traditional MIA-QSAR is recommended rather than the augmented approaches when the chemical sampling are not trivial to be constructed (drawn) using the available programs (*e.g.* very complex tridimensional geometries), such as GaussView, or when atomic sizes and electronegativity are not expected to play a determinant role to improve the model quality, *e.g.* for hydrocarbon chains only as substituents in a congeneric series. Similarly, aug-MIA-QSAR$_{color}$ is expected to give better outcomes than aug-MIA-QSAR if electronegativity is indeed an effective descriptor and if the series contains structural dependence with atomic electronegativity. For instance, a series of compounds containing chlorine and nitrogen in different positions of an aromatic ring as substituents does not need to be modeled with aug-MIA-QSAR$_{color}$, since both atoms have Pauling electronegativity 3.0.

REFERENCES

Antunes, J. E., Freitas, M. P., da Cunha, E. F. F., Ramalho, T. C., & Rittner, R. (2008). In silico prediction of novel phosphodiesterase type-5 inhibitors derived from Sildenafil, Vardenafil and Tadalafil. *Bioorganic & Medicinal Chemistry*, *16*(16), 7599–7606. doi:10.1016/j.bmc.2008.07.022 PMID:18656371

Bitencourt, M., & Freitas, M. P. (2008). MIA-QSAR evaluation of a series of sulfonylurea herbicides. *Pest Management Science*, *64*(8), 800–807. doi:10.1002/ps.1565 PMID:18338340

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Bro, R. (1996). Multiway calibration: Multilinear PLS. *Journal of Chemometrics*, *10*(1), 47–61. doi:10.1002/(SICI)1099-128X(199601)10:1<47::AID-CEM400>3.0.CO;2-C

Bro, R. (1997). PARAFAC: Tutorial and applications. *Chemometrics and Intelligent Laboratory Systems*, *38*(2), 149–171. doi:10.1016/S0169-7439(97)00032-4

Brown, R. D., & Martin, Y. C. (1997). The information content of 2D and 3D structural descriptors relevant to ligand-receptor binding. *Journal of Chemical Information and Computer Sciences*, *37*(1), 1–9. doi:10.1021/ci960373c

Chirico, N., & Gramatica, P. (2012). Real external predictivity of QSAR models. Part 2. New intercomparable thresholds for different validation criteria and the need for scatter plot inspection. *Journal of Chemical Information and Modeling*, *52*(8), 2044–2058. doi:10.1021/ci300084j PMID:22721530

Cormanich, R., Nunes, C. A., & Freitas, M. P. (2012). Chemical drawings correlate to biological properties: MIA-QSAR. *Quimica Nova*, *35*, 1157–1163. doi:10.1590/S0100-40422012000600017

Cormanich, R. A., Goodarzi, M., & Freitas, M. P. (2009). Improvement of multivariate image analysis applied to quantitative structure–activity relationship (QSAR) analysis by using wavelet-principal component analysis ranking variable selection and least-squares support vector machine regression: QSAR study of checkpoint kinase WEE1 inhibitors. *Chemical Biology & Drug Design*, *73*(2), 244–252. doi:10.1111/j.1747-0285.2008.00764.x PMID:19207427

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

Crum-Brown, A., & Fraser, T. R. (1869). On the connection between chemical constitution and physiological action. Part 1: On the physiological action of the ammonium bases derived from Strychia, Brucia, Thebia, Codeia, Morphia, and Nicotia. *Transactions of the Royal Society of Edinburgh*, *25*, 257–274.

Davis, L. (Ed.). (1991). *Handbook of genetic algorithms*. London: Van Nostrand-Reinhold.

Deeb, O., Alfalah, S., Freitas, M. P., da Cunha, E. F. F., & Ramalho, T. C. (2012). Exploring MIA-QSARs for farnesyltransferase inhibitory effect of antimalarial compounds refined by docking simulations. *Journal of Biophysical Chemsitry*, *3*(01), 58–71. doi:10.4236/jbpc.2012.31008

Dennington, R. D., II, Keith, T. A., & Millam, J. M. (2008). GaussView 5.0. Wallingford.

Du, X., Guo, C., Hansell, E., Doyle, P. S., Caffrey, C. R., & Holler, T. P. et al. (2002). Synthesis and structure-activity relationship study of potent trypanocidal thio semicarbazone inhibitors of the trypanosomal cysteine protease cruzain. *Journal of Medicinal Chemistry*, *45*(13), 2695–2707. doi:10.1021/jm010459j PMID:12061873

Estrada, E., Molina, E., & Perdomo-López, I. (2001). Can 3D structural parameters be predicted from 2D (topological) molecular descriptors? *Journal of Chemical Information and Computer Sciences*, *41*(4), 1015–1021. doi:10.1021/ci000170v PMID:11500118

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

Freitas, M. P. (2006). MIA-QSAR modelling of anti-HIV-1 activities of some 2-amino-6-arylsulfonylbenzonitriles and their thio and sulfinyl congeners. *Organic & Biomolecular Chemistry*, *4*(6), 1154–1159. doi:10.1039/b516396j PMID:16525561

Freitas, M. P., Brown, S. D., & Martins, J. A. (2005). MIA-QSAR: A simple 2D image-based approach for quantitative structure-activity relationship analysis. *Journal of Molecular Structure*, *738*(1-3), 149–154. doi:10.1016/j.molstruc.2004.11.065

Freitas, M. P., da Cunha, E. F. F., Ramalho, T. C., & Goodarzi, M. (2008). Multimode methods applied on MIA descriptors in QSAR. *Current Computer-aided Drug Design*, *4*(4), 273–282. doi:10.2174/157340908786786038

Freitas, M. P., & Ramalho, T. C. (2013). Employing conformational analysis in the molecular modeling of agrochemicals: Insights on QSAR parameters of 2,4-D. *Ciência & Agrotecnologia*, *37*(6), 485–494. doi:10.1590/S1413-70542013000600001

Freitas, M. P., & Rittner, R. (2008). MIA-QSAR as an alternative approach for modeling some antifungals. *QSAR & Combinatorial Science*, *27*(5), 582–585. doi:10.1002/qsar.200710047

Garkani-Nejad, Z., & Ahmadi-Roudi, B. (2010). Modeling the antileishmanial activity screening of 5-nitro-2-heterocyclic benzylidene hydrazides using different chemometrics methods. *European Journal of Medicinal Chemistry*, *45*(2), 719–726. doi:10.1016/j.ejmech.2009.11.019 PMID:19959260

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

Goodarzi, M., da Cunha, E. F. F., Freitas, M. P., & Ramalho, T. C. (2010). QSAR and docking studies of novel antileishmanial diaryl sulfides and sulfonamides. *European Journal of Medicinal Chemistry*, *45*(11), 4879–4889. doi:10.1016/j.ejmech.2010.07.060 PMID:20728249

Goodarzi, M., & Freitas, M. P. (2009). Prediction of electrophoretic enantioseparation of aromatic amino acids/esters through MIA-QSPR. *Separation and Purification Technology*, *68*(3), 363–366. doi:10.1016/j.seppur.2009.06.005

Goodarzi, M., Freitas, M. P., & Ferreira, E. B. (2009). Influence of changes in 2-D chemical structure drawings and image formats on the prediction of biological properties using MIA-QSAR. *QSAR & Combinatorial Science*, *28*(4), 458–464. doi:10.1002/qsar.200810146

Goodarzi, M., Freitas, M. P., & Jensen, R. (2009). Ant colony optimization as a feature selection method in the QSAR modeling of anti-HIV-1 activities of 3-(3,5-dimethylbenzyl)uracil derivatives using MLR, PLS and SVM regressions. *Chemometrics and Intelligent Laboratory Systems*, *98*(2), 123–129. doi:10.1016/j.chemolab.2009.05.005

Goodarzi, M., Freitas, M. P., & Jensen, R. (2009). Feature selection and linear/nonlinear regression methods for the accurate prediction of glycogen synthase kinase-3β inhibitory activities. *Journal of Chemical Information and Modeling*, *49*(4), 824–832. doi:10.1021/ci9000103 PMID:19338295

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Guimarães, M. C., Silva, D. G., da Mota, E. G., da Cunha, E. F. F., & Freitas, M. P. (2014). Computer-assisted design of dual-target anti-HIV-1 compounds. *Medicinal Chemistry Research*, *23*(3), 1548–1558. doi:10.1007/s00044-013-0765-3

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). 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

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 Medicinal Chemistry*, *119*, 10509–10524.

Jana, G. A., Delgado, E. J., & Medina, F. E. (2014). How important is the synclinal conformation of sulfonylureas to explain the inhibition of AHAS: A theoretical study. *Journal of Chemical Information and Modeling*, *54*(3), 926–932. doi:10.1021/ci400721y PMID:24548139

Jensen, R. (2006). *Studies in computational intelligence* (A. Abraham, C. Grosan, & V. Ramos, Eds.). Heidelberg: Springer.

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*, *23*(1-3), 4–25. doi:10.1016/S0169-409X(96)00423-1 PMID:11259830

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

Martin, Y. C. (2009). Let's not forget tautomers. *Journal of Computer-Aided Molecular Design*, *23*(10), 693–704. doi:10.1007/s10822-009-9303-2 PMID:19842045

Meyer, H. H. (1899). On the theory of alcohol narcosis: First communication. Which property determines its narcotic effect? *Archives of Experimental Pathology and Pharmacology*, *42*, 109–118. doi:10.1007/BF01834479

Mitra, I., Saha, A., & Roy, K. (2010). Exploring quantitative structure–activity relationship studies of antioxidant phenolic compounds obtained from traditional Chinese medicinal plants. *Molecular Simulation*, *36*(13), 1067–1079. doi:10.1080/08927022.2010.503326

Nunes, C. A., & Freitas, M. P. (2013). Introducing new dimensions in MIA-QSAR: A case for chemokine receptor inhibitors. *European Journal of Medicinal Chemistry*, *62*, 297–300. doi:10.1016/j.ejmech.2013.01.005 PMID:23357311

Ojha, P. K., Mitra, I., Das, R. N., & Roy, K. (2011). Further exploring r^2_m metrics for validation of QSPR models. *Chemometrics and Intelligent Laboratory Systems*, *107*(1), 194–205. doi:10.1016/j.chemolab.2011.03.011

Overton, E. (1901). *Studien Uber die Narkose*. Jena, Germany: Fischer.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Pan, D., Tseng, Y., & Hopfinger, A. J. (2003). Quantitative structure-based design: Formalism and application of receptor-dependent RD-4D-QSAR analysis to a set of glucose analogue inhibitors of glycogen phosphorylase. *Journal of Chemical Information and Computer Sciences*, *43*(5), 1591–1607. doi:10.1021/ci0340714 PMID:14502494

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

Ramalho, T. C., Freitas, M. P., & da Cunha, E. F. F. (Eds.). (2012). *Chemoinformatics: Directions toward combating neglected diseases*. Dubai: Bentham.

Richet, M. C. (1893). Sur le rapport entre la toxicité et les propriétés physiques des corps. *Comptes Rendus des Seances de la Société de Biologie et de ses Filiales*, *45*, 775–776.

Roy, K., Chakraborty, P., Mitra, I., Ojha, P. K., Kar, S., & Das, R. N. (2013). Some case studies on application of "r_m^2" 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

Roy, K., Mitra, I., Kar, S., Ojha, P. K., Das, R. N., & Kabir, H. (2012). Comparative studies on some metrics for external validation of QSPR models. *Journal of Chemical Information and Modeling*, *52*(2), 396–408. doi:10.1021/ci200520g PMID:22201416

Roy, P. P., Paul, S., Mitra, I., & Roy, K. (2009). On two novel parameters for validation of predictive QSAR models. *Molecules (Basel, Switzerland)*, *14*(5), 1660–1701. doi:10.3390/molecules14051660 PMID:19471190

Selassie, C. D. (2003). *Burger's medicinal chemistry and drug discovery* (D. J. Abraham, Ed.). New York: John Wiley & Sons.

Silva, D. G., Freitas, M. P., da Cunha, E. F. F., Ramalho, T. C., & Nunes, C. A. (2012). Rational design of small modified peptides as ACE inhibitors. *Medicinal Chemistry Communications*, *3*(10), 1290–1293. doi:10.1039/c2md20214j

Teófilo, R. F., Martins, J. P. A., & Ferreira, M. M. C. (2009). Sorting variables by using informative vectors as a strategy for feature selection in multivariate regression. *Journal of Chemometrics*, *23*(1), 32–48. doi:10.1002/cem.1192

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., & 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

Whitley, D. (1994). A genetic algorithm tutorial. *Statistics and Computing*, *4*(2), 65–85. doi:10.1007/BF00175354

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Zingales, B., Andrade, S. G., Briones, M. R. S., Campbell, D. A., Chiari, E., & Fernandes, O. et al. (2009). A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: Second revision meeting recommends TcI to TcVI. *Memorias do Instituto Oswaldo Cruz*, *104*(7), 1051–1054. doi:10.1590/S0074-02762009000700021 PMID:20027478

ADDITIONAL READING

Charton, M. (1999). *Advances in Molecular Structure Research*. Greenwich: JAI Press.

Damale, M. G., Harke, S. N., Khan, F. A. K., 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:10.2174/13895575113136660104 PMID:24195665

Esbensen, K., & Geladi, P. (1989). Strategy of multivariate image analysis (MIA). *Chemometrics and Intelligent Laboratory Systems*, *7*(1-2), 67–86. doi:10.1016/0169-7439(89)80112-1

Esbensen, K. H., Hjelmen, K. H., & Kvaal, K. (1996). The AMT approach in chemometrics - first forays. *Journal of Chemometrics*, *10*(5-6), 569–590. doi:10.1002/(SICI)1099-128X(199609)10:5/6<569::AID-CEM466>3.0.CO;2-W

Ferreira, M. M. C. (2002). Multivariate QSAR. *Journal of the Brazilian Chemical Society*, *13*(6), 742–753. doi:10.1590/S0103-50532002000600004

Freitas, M. P. (2007). Multivariate QSAR: From classical descriptors to new perspectives. *Current Computer-aided Drug Design*, *3*(4), 235–239. doi:10.2174/157340907782799408

Freitas, M. P. (2009). MIA-QSTR study of different organic compounds to *Pimephales promelas. Medicinal Chemistry Research*, *18*(8), 648–655. doi:10.1007/s00044-008-9156-6

Freitas, M. R., Matias, S. V. B. G., Macedo, R. L., Freitas, M. P., & Venturin, N. (2013). Augmented multivariate image analysis applied to quantitative structure−activity relationship modeling of the phytotoxicities of benzoxazinone herbicides and related compounds on problematic weeds. *Journal of Agricultural and Food Chemistry*, *61*(36), 8499–8503. doi:10.1021/jf4024257 PMID:23947385

Geladi, P., & Esbensen, K. (1991). Regression on multivariate images: Principal component regression for modeling, prediction and visual diagnostic tools. *Journal of Chemometrics*, *5*(2), 97–111. doi:10.1002/cem.1180050206

Geladi, P., & Grahn, H. (1996). *Multivariate Image Analysis*. Chichester: Wiley.

Geladi, P., & Kowalski, B. R. (1986). Partial least-squares regression: A tutorial. *Analytica Chimica Acta*, *185*, 1–17. doi:10.1016/0003-2670(86)80028-9

Geladi, P., Wold, S., & Esbensen, K. (1986). Image analysis and chemical information in images. *Analytica Chimica Acta*, *191*, 473–480. doi:10.1016/S0003-2670(00)86335-7

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Goodarzi, M., & Freitas, M. P. (2008). Predicting boiling points of aliphatic alcohols through multivariate image analysis applied to quantitative structure-property relationships. *The Journal of Physical Chemistry A*, *112*(44), 11263–11265. doi:10.1021/jp8059085 PMID:18841953

Goodarzi, M., Freitas, M. P., & Ramalho, T. C. (2009). Prediction of ^{13}C chemical shifts in methoxyflavonol derivatives using MIA-QSPR. *Spectrochimica Acta. Part A: Molecular Spectroscopy*, *74*(2), 563–568. doi:10.1016/j.saa.2009.07.003 PMID:19648055

Guimarães, M. C., da Mota, E. G., Silva, D. G., & Freitas, M. P. (2014). aug-MIA-QSPR modelling of the toxicities of anilines and phenols to *Vibrio fischeri* and *Pseudokirchneriella subcapitata. Chemometrics and Intelligent Laboratory Systems*, *134*, 53–57. doi:10.1016/j.chemolab.2014.03.005

Hansch, C., & Leo, A. (1979). *Substituent constants for correlation analysis in Chemistry and Biology*. New York: Wiley.

Hansch, C., Leo, A., & Hoekman, D. (1979). *Exploring QSAR. Hydrophobic, electronic, and steric constants*. Washington: American Chemical Society.

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

Maya, J. D., Cassels, B. K., Iturriaga-Vásquez, P., Ferreira, J., Faúndez, M., & Galanti, N. et al. (2007). Mode of action of natural and synthetic drugs against *Trypanosoma cruzi* and their interaction with the mammalian host. *Comparative Biochemistry and Physiology. Part A, Molecular & Integrative Physiology*, *146*(4), 601–620. doi:10.1016/j.cbpa.2006.03.004 PMID:16626984

Shahlaei, M. (2013). Descriptor selection methods in quantitative structure-activity relationship studies: A review study. *Chemical Reviews*, *113*(10), 8093–8103. doi:10.1021/cr3004339 PMID:23822589

Silla, J. M., Nunes, C. A., Cormanich, R. A., Guerreiro, M. C., Ramalho, T. C., & Freitas, M. P. (2011). MIA-QSPR and effect of variable selection on the modeling of kinetic parameters related to activities of modified peptides against dengue type 2. *Chemometrics and Intelligent Laboratory Systems*, *108*(2), 146–149. doi:10.1016/j.chemolab.2011.06.009

Todeschini, R., Consonni, V., Mannhold, R., Kubinyi, H., & Timmerman, H. (2008). *Handbook of Molecular Descriptors*. Weinheim: Wiley-VCH.

Tropsha, A. (2010). Best practices for QSAR model development, validation, and exploration. *Molecular informatics*, 29, 476-488.

Wold, H. (1966). *Multivariate Analysis* (K. R. Krishnaiah, Ed.). New York: Academic Press.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

KEY TERMS AND DEFINITIONS

log*P*: The logarithm of the octanol/water partition coefficient. It measures the hydrobobicity of a substance and was the milestone parameter used in QSAR.

MIA: Multivariate image analysis. First explored by Paul Geladi (University of Umea) in hyperspectral remote sensing analysis and further used in chemical applications.

PCA: Principal component analysis, a multivariate method widely used for pattern recongnition. It is mathematically defined as an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by some projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on.

PLS: Partial least squares, a multivariate method widely used for linear regression. PLS is used to find the fundamental relations between two matrices (**X** and **Y**), i.e. a latent variable approach to modeling the covariance structures in these two spaces.

QSAR: Quantitative structure-activity relationship. QSAR regression models relate a set of "predictor" variables (**X**) to the potency of the response variable (**Y**), while classification QSAR models relate the predictor variables to a categorical value of the response variable.

Trypanosomiasis: The name of several diseases invertebrates caused by parasitic protozoan trypanosomes of the genus *Trypanosoma*. In the Latin America, the human form of trypanosomiasis is called Chagas disease and it is cause by the protozoan *T. cruzi*.

Validation: The main procedure in QSAR to attest its reliability and predictability. The most common validation methods include cross-validation and external validation.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 4
Quantitative Structure–Activity/Property/Toxicity Relationships through Conceptual Density Functional Theory–Based Reactivity Descriptors

Sudip Pan
Indian Institute of Technology Kharagpur, India

Venkatesan Subramanian
Central Leather Research Institute, India

Ashutosh Gupta
Udai Pratap Autonomous College, India

Pratim K. Chattaraj
Indian Institute of Technology Kharagpur, India

ABSTRACT

Developing effective structure-activity/property/toxicity relationships (QSAR/QSPR/QSTR) is very helpful in predicting biological activity, property, and toxicity of a given set of molecules. Regular change in these properties with the structural alteration is the main reason to obtain QSAR/QSPR/QSTR models. The advancement in making different QSAR/QSPR/QSTR models to describe activity, property, and toxicity of various groups of molecules is reviewed in this chapter. The successful implementation of Conceptual Density Functional Theory (CDFT)-based global as well as local reactivity descriptors in modeling effective QSAR/QSPR/QSTR is highlighted.

1. INTRODUCTION

The properties of molecules mainly depend on the distribution of electron density within the individual molecules. It is, therefore, believed that molecules having similar structures will exhibit similar properties. On this hypothesis, the idea of structure-activity relationship (SAR) exists (Nantasenamat et al., 2009). Accordingly, it is assumed that molecules with similar structures exhibit comparable activities (properties), and, so on the basis of known properties of a set of molecules, prediction of the properties of unknown molecules having similar structures could be made, provided appropriate models are de-

DOI: 10.4018/978-1-4666-8136-1.ch004

Copyright © 2015, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

veloped. Models which are able to predict the properties of an unknown molecule on a qualitative basis are termed as Qualitative Structure-Activity Relationship models (Arning et al., 2008), and, the ones which can quantify such relationship are termed as Quantitative Structure-Activity Relationship (QSAR) models (Gramatica, 2007). Such QSAR models are very useful in the design of molecules which can serve as a better drug or in the manufacturing of materials which could serve as better substance in terms of utility, economy, durability and environmental effects in totality. The quest for such molecules begins with SAR studies. On the basis of activities of a set of molecules of similar structures or parameters, different SAR models are designed. Thereafter, those molecules which are found to have a potential to show a desired characteristic (property/activity) are filtered with the aid of such models, and, then the systematic efforts towards their syntheses begin. Thus, such an approach helps in the reduction of the cost effectively. Since a molecule can have different kinds of properties, a large number of QSAR models could be developed for the same set of molecules depending on their properties. Such properties could be physico-chemical properties, quantum-chemical properties or even toxicity properties. It clearly implies that not only the unknown toxicity parameter of a molecule of whose different parameters are yet to be assessed, could be determined, but its other set of properties could also be determined provided effective QSAR models are generated. When the attempt is to determine the toxicological effect of a molecule based on a model having correlation between structure and toxicity of already determined molecules, then such an approach is defined as Quantitative Structure-Toxicity Relationship (QSTR) model (Roy and Ghosh, 2004). Such models are very effective in the prediction of toxicity of molecules which have potential to act as a toxic substance. Thus, QSTR helps in building up of an effective regression model that helps in predicting the reactivity of an unknown molecule on the basis of behavior of an analogous set of chemically or structurally similar compounds.

QSTR models are proposed based on different parameters. Most often these parameters are defined in terms of descriptors. These descriptors could be based on different experimental and theoretical properties such as physico-chemical and quantum-chemical properties, respectively. One of the most popular and cheap methods in the field of quantum-chemical prediction of the properties of molecules is Density Functional Theory (DFT) (Parr & Yang, 1989). It predicts the properties of a molecule based on its molecular electron density. A closely related method describing the chemical properties such as electronegativity, ionization potential, hardness etc with the aid of DFT method is known as Conceptual Density Functional Theory method (CDFT) (Chattaraj, 2009; Chakraborty et al., 2010; Chattaraj & Giri, 2009; Geerlings et al., 2003). This method has been successfully and exhaustively employed in the generation of different descriptors, which have been able to produce QSAR and QSTR models of very high efficiency. Thus, CDFT has helped in the assessment of activity and toxicity of different molecules within the ambit of QSAR and QSTR respectively. It has been found that CDFT is quite cheaper in comparison to other expensive quantum-chemical methods such as *ab initio* methods. CDFT has been proven to develop effective and efficient QSTR models.

In this chapter, at first we have provided a general introduction highlighting the necessity of structure-activity/toxicity based regression models. Various QSAR/QSPR/QSTR based methods are outlined in section 2. Section 3 describes the conceptual density functional theory and the related global and local reactivity descriptors. Section 4 deals with the utility of electrophilicity index and net electrophilicity index and their local variants in predicting toxicity of various classes of molecules. Section 5 mentions about the importance of the group philicity towards modeling an effective QSAR. Section 6 highlights the effect of studied levels and / or basis sets in constructing QSAR/QSPR/QSTR models. Section 7 concludes the chapter by mentioning the effectiveness of various CDFT based descriptors in building QSAR/QSPR/QSTR models.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

2. QUANTITATIVE STRUCTURE-ACTIVITY/ PROPERTY/TOXICITY RELATIONSHIPS

Quantitative Structure Activity Relationship (QSAR) acts as a function which connects structure of a molecule to different properties. Those properties which show medicinal applications are considered as an activity of a molecule, and, are thus studied within the purview of QSAR. In the present world of global recession, such a technique is able to save money and time. Much of the efforts are now directed towards development of models which predict accurately the properties of a molecule on the basis of their structures. Thus, QSAR models are able to predict biological activities and physico-chemical properties and rationalize the mechanism of action for different chemical reactions. Normally, in order to develop a QSAR, five to ten compounds are necessary for each descriptor as argued by Topliss and Costello (1972) and further mentioned by Schultz et al. (2003c). However, in order to develop robust QSAR, many more compounds are required. With the advent of sophisticated techniques, new software, and, hardware which works on novel algorithms, has made the task of generation of an accurate model within the purview of QSAR simplier. QSAR has its genesis in the pioneering work of Hansch and coworkers (1962, 1964) wherein they used linear multiple regression models to describe the biological property of the compounds, which were not synthesized till then. Different parameters such as hydrophobicity, electronic effects and steric properties were utilized to generate a relationship among them and the concentration of the molecule. Free and Wilson (1964) and Fujita et al. (1971) developed non-parameteric methods to define biological activity. Later Linear Solvation Free Energy Relationship (LSFER) was proposed which predicted the properties of a molecule on solute-solvent interactions (Kamlet et al., 1983; Taft et al., 1985). It consists of different solvatochromic descriptors such as cavity/bulk term, polarizability term, and H-bond terms which are used to derive a relationship between overall property of a molecule and different parameters. A good account of the role of QSAR in chemical toxicology was elaborated by Schultz et al. (2003a, 2003b). Other articles also highlight the development of the QSAR methodology and their role in the prediction of bio-toxicity (Gombar et al., 2006; Kar & Roy, 2010; Nantasenamat et al., 2009; Puzyn et al., 2010; Selassie et al., 2010). Hammett (1935, 1937) by employing correlation principles, in 1930s had efficiently shown the effect of structural changes on reaction mechanism of substituted benzoic acid derivatives. It led subsequently to the birth of well known linear free-energy relationship essentially by connecting Gibbs free energy of reaction and Gibbs free energy of activation which in a broader sense established a connection between thermodynamics and kinetics. The concept was further extended by Taft (1956) for aliphatic compounds. This paved the way of strong QSAR models as was evident from the further work carried out by Hansch and coworkers (1993, 1996) wherein they discussed the role of QSAR. The role of QSAR in the study of chemico-biological interactions (Blaney & Hansch, 1990; Debnath et al., 1994; Gao et al., 1999; Hansch et al., 1986, 1989, 1995, 1997, 1998, 2001; Hansen et al., 1995; Langridge & Klein, 1990; Lien et al., 1968;) and also for the drug design (Bajot, 2006; Garg et al., 1999; Hansch et al., 1973, 1995; Kubinyi, 1997; Maddalena, 1998; Norinder, 1998; Voskresensky & Levitsky, 2002; Yang & Huang, 2006;) was also established.

In order to generate an effective QSAR model, the first and foremost requirement is the collection of data and generation of database. This helps in defining training sets which generally consist of known data obtained from various sources. With the aid of such data, different kinds of descriptors are developed on the basis of molecular formula (1-D descriptor), the 2-dimensional structure (2-D descriptor), and the 3-dimensional conformation (3-D descriptor) of a molecule. Generally, molecular descriptors generated from some special types of atoms or fragments counting or based on bulk and physico-chemical proper-

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

ties like molecular weight, atomic number, number of OH groups, number of non-hydrogenic atoms and so on correspond to 1-D descriptors.2-D descriptors are usually obtained by applying algorithms to the topology based representation. 3-D descriptors, also called geometrical descriptors are generated from the spatial coordinates (x,y,z) of a molecular system whereas the interaction energy values between an embedded molecule in a grid and a probe will provide 4-D descriptors. In general, molecular descriptors should satisfy some basic requirements as suggested by Randić (1991). Different other articles also showed the role of QSAR in the generation of different algorithms (Karcher & Devillers, 1990; Roy & Mitra, 2009; Selassie et al., 2002).

Molecular descriptors are mathematical representations of various properties of a molecule which are generally derived after the application of a reliable algorithm to the molecular representation or are determined from experiment. Being related to the molecular Hamiltonian, the molecular descriptors like electron density, electrostatic potential, etc. would represent all the properties of a molecule (Bonaccorsi et al., 1970; Murray & Politzer, 2003; Pathak & Gadre, 1990; Politzer, 1988). These descriptors help in the development of models for chemistry, toxicology, health research, pharmaceutical sciences, etc. They are generated by the application of different principles from different fields and theory, such as organic chemistry, graph theory, quantum-chemistry, information theory etc. The electron localization and delocalization functions obtained from the quantum theory of atoms in molecules (QTAIM) are also very effective in modeling QSAR (Buttingsrud et al., 2007a, 2007b, 2007c; Chaudry & Popelier, 2003, 2004; O'Brien & Popelier, 1999; Popelier et al., 1999, 2004; Smith & Popelier, 2005). Popelier et al. (1999) showed that a QTAIM based approach called quantum topological molecular similarity (QTMS) could predict the toxicity of health-hazardous lipid-soluble environmental pollutants as efficient as that of comparative molecular field analysis (CoMFA). Very recently, Matta et al. (2014) reviewed the applicability of the electron density, electron localization and delocalization functions, and the electrostatic potential in developing effective QSAR models.

Bohari et al. (2011) adopted so-called analogue-based approach to predict the anti-cancer activity of 266 different compounds against 29 various cancer cell lines. Various models were constructed through 1 to 10 different descriptors. They found that in general three-parameters regression models could predict the anti-cancer property satisfactorily. Filho et al. (2012) performed QSAR studies on 32 Morita–Baylis–Hillman adducts against Leishmania amazonensis. Energy of the lowest unoccupied Kohn–Sham orbital, number of $o-NO_2$ group, number of $-CH_2-$ group and an increment in log P parameters were used to build the regression model. A reasonably good predictivity was found for this model. QSAR studies were also performed in two sets of antitumour drugs 2-(4-aminophenyl)benzothiazoles by using different indices encompassing constitutional, geometrical, topological, and electronic descriptors (Hilal & Elroby, 2011). Good correlation having a range of r^2 as 0.867-0.954 was attained between computed and experimental activities with these models. A number of molecular descriptors like polar surface area, Randic shape index, number of phenolic groups and energy of the highest occupied molecular orbital were used to predict the estrogenic activity of different terpenoid ester derivatives obtained from *Ferula* plants (Rasulev et al., 2007). An r^2 value of 0.892 implied the reliability of the regression model to predict the activity of these compounds. Li et al. (2010) carried out the QSAR studies on 27 diacyl-hydrazine derivatives by using DFT based method and semi-empirical AM1-MOPAC method. They found higher predictivity at the DFT level than that at the semi-empirical level. 3D-QSAR models were also generated based on the geometries obtained at the DFT level and electrostatic potential fitting charges. The results of the QSAR and 3D-QSAR were found to be comparable. For these compounds, it was argued that the electrostatic and hydrophobic properties are important in deciding their overall biological activities.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Zhang et al. (2009) combined DFT-based conformation analysis with QSAR to develop a simple new DFT/QSAR approach to predict the nature of bioactive conformation in the side chain of cyclic imide derivatives as protoporphyrinogen oxidase inhibitors. The potential energy surface scan, molecular docking and molecular dynamic simulation supported the reliability of this new approach. Damme and Bultinck (2010) employed conceptual DFT bsed quantum chemical molecular field in order to construct 3D-QSAR models to describe the antituberculotic activity of salicylamide derivatives. QSAR studies on the interactions between the terpenoid mosquito repellents and lactic acid were made by Song et al. (2013). They found that in addition to the structures of the repellents, the repellent–lactic acid complexes also have a role in deciding their activities. Use of ordered predictors selection algorithm and partial least squares regression was also made to select variables and to construct QSAR models of various 4,5-dihydroxypyrimidine carboxamide derivatives as HIV-1 integrase inhibitors (de Melo et al., 2009). The electronic distribution of the studied compounds appeared to have deciding role in their inhibition activities. The antifungal activity of N-nitrosopiperidone semicarbazone derivatives against the plant pathogens viz., *Fusarium oxysporum* and *Rhizoctonia solani* can also be predicted by the regression models devepoled by the descriptors like molecular electrostatic potential, charge on N center, dipole moment, minium surface area and log P (Hemalatha et al., 2012). The QSAR studies on the imidazole and benz-imidazole derivatives as corrosion inhibitors revealed that HOMO energy, polarizability, dipole, frontier orbital charge density, the interaction mode between inhibitors and metal surface and steric hindrance of molecules are deciding descriptors to represent their inhibiting potency (Zhang et al., 2005). Various DFT based descriptors like HOMO-LUMO energy gap, HOMO and LUMO energies, mean molecular polarizability, atomic charge, weighted nucleophilic atomic frontier electron density were employed in order to build regression models for the various cyclic imide derivatives as protoporphyrinogen oxidase (Protox) inhibitors (Wan et al., 2004). The DFT-based QSAR and 3D-QSAR approaches on the activity of aurone derivatives against chloroquine resistant parasite by Adhikari et al. (2013) are also notworthy. A review summarizing various quantum-chemical descriptors in developing QSAR and QSPR written by Karelson and Lobanov (1996) is highly recommended to get more details about these approaches.

Geometrical descriptors were designed and included in the QSAR during the development of 3D-QSAR models in mid 1980s. These included gravitational indices (Katritzky et al., 1996), shadow indices (Rohrbaugh & Jurs, 1987), EVA descriptors (Ferguson et al., 1997), GETAWAY descriptors (Consonni et al., 2002), 3D-Morse descriptors (Schuur et al., 1996), charged partial surface area descriptors (Stanton & Jurs, 1990), and WHIM descriptors (Todeschini et al., 1994).

The descriptors such as topological descriptors utilize molecular graph. Different topological indices such as those developed by Weiner (1947), Randić (1975), Balaban (1998), Kier and Hall (1995) were reported in literature to generate QSAR models. Other descriptors such as steric (Camenisch et al., 1998; Mezey, 1992) viz., volume, shape and surface, were found to be useful in representing the global structural characteristics of a molecule and different software such as VolSurf software (Cruciani et al., 2000) were employed to create effective QSAR models. In order to generate 2D and 3D descriptors for the prediction of physicochemical properties, the methods provided by Rekker (1992) were used.

CoMFA (Cramer et al., 1988) is an effective method in building effective 3D-QSAR, which was successfully applied in drug designing and other related applications (Aboye et al., 2004; Akamatsu, 2002; Amin & Welsh, 2006; Avery et al., 2002; Bhongade & Gadad, 2004; Hannongbua etal., 2001; Jayatilleke et al., 2000; Sippl & Holtje, 2000). History behind the idea of CoMFA in modeling QSAR is that in the classical QSAR, the different biological acitivies of drugs or other compounds of interest are correlated with various physico-chemical properties or quantum-chemical descriptors, provided

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

they could represent certain structural characteristics. Therefore, they did not take care of the information regarding their 3-D geometries and chirality. On the other hand, at longer distance an active site of a drug can only experience the electrostatic field of a target molecule and at shorter distance its hard body including the charge distribution (Kubinyi, 1998). Other two important observations were that the most of the molecular interactions which generate biological activity, are of non-covalent type and the most of the molecular mechanics force fields, wich describe the non-covalent interaction as steric and electrostatic forces, could explain a large variety of molecular properties. Therefore, it appeared that an appropriate sampling of the steric and electrostatic forces surrounding the considered molecules could give all the necessary information regarding their observed biological properties. Cramer and Milne (1979) compared the molecules for the first time by arranging them in space and by molecular field mapping in a 3-D grid. The same approach was then implemented in dynamic lattice-oriented molecular modeling system (DYLOMMS) method (Kubinyi, 1993). However, to obtain success in broader applications of this approach, certain factors like the use of partial least square analysis instead of principal component analysis and use of appropriate software are important (Kubinyi, 1998). In 1988, Cramer et al. (1988) termed this approach as CoMFA. The major precondition for this approach is that all the selected molecules under study must interact with the receptors of same kind and in same mechanistic way. Then in order to develop a CoMFA model, a traning set is chosen from a certain subgroup of the considered molecules and the remaining part is treated as a test set. This test set is used to verify the efficacy of this model to describe certain properties. The final output of this analysis provides a regression equation having several coefficients and they are very often depicted as a set of contour plots. From these plots, the favorable and unfavorable steric regions and favorable and unfavorable regions for electropositive or electronegative substituents could be identified. Then one could predict the desired properties of the test set or other related samples by qualitative examination of these plots or by computing the fields of these molecules quantitatively. Cramer et al. (1988) first employed this method to predict the binding affinities of steroids towards the human corticosteroid. The squared correlation coefficient (r^2) for the affinity values was found to be very high (close to 0.9). Since 1988, a lot of articles including reviews are reported in the literature on this topic (Adhikari et al., 2014; Blankley, 1996; Kim, 1995; Martin et al., 1996; Martin & Lin, 1996).

In addition to the development of the QSAR, quantitative structure-property relationship (QSPR) was also used to predict a certain property of the molecules. The hypothesis, that structural alteration changes the observed macroscopic properties of systems is the base of QSPR methods. Several research groups contributed significantly in the development of QSPR models (Abraham, 1994; Abraham et al., 1995; Balaban, 1997; Hilal et al., 1994; Katritzky et al., 1995; Kier & Hall, 1986; Lucic & Trinajstic, 1999; Murray & Politzer, 1994; Randic & Razinger, 1996; Stuper et al., 1979). The efficacy of QSPR as a useful tool in connecting molecular structure with the various physical, chemical, biological and technological properties was further demonstrated by Katritzky et al. (2000). The successful implication of QSPR in medicinal chemistry to predict different property was also available in the literature (Katritzky et al., 2002). A fair correletion between structure and property was obtained by using QSPR models made on an empirical basis even for large data sets. The importance of data reduction methods like principal component analysis (PCA) of a matrix generated by the different properties of the systems was also found in providing insight into these properties quantitatively (Leo & Hansch, 1999). The reviews summarizing the construction of QSPR models in predicting diverse properties of materials are highly recommended. Variuos approachs were also made to model effective QSPR for anti HIV drugs. The applicability of QSPR in pharmaceutical arena including different pharmacokinetic aspects towards drug designing was reviewed by Grover et al. (2000a, 2000b).

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Further, quantitative structure-toxicity relationship (QSTR) is essential in order to have proper knowledge about the toxicity of various essential systems; hence it is an issue of environmental safety and security. The approach of QSTR modeling even becomes more popular since experimental determination of toxicity of each and every essential compound is hard with respect to time, cost, and the availability of resources. Different theoretical approaches have been made to connect the structural features (electronic, geometric, topological, quantum chemical, etc.) with the toxicity. An effective QSTR model could predict the relative toxicity/safety of different systems in a reliable way saving the experimental effort. Cronin and Schultz (1998) developed useful QSTR model for v*ibrio pscheri* toxicity including all probable mechanisms. Schultz et al. (1999, 2002) evaluated toxicity of numerous aliphatic chemicals and substituted benzenes towards *Tetrahymena pyriformis*.

The databases regarding toxicity of a number of fish species like the guppy (*Poecilia reticulata*) and zebrafish (*Brachydanio rerio*) are also available. The fish data sets for the fathead minnow (*Pimephales promelas*) to predict the acute toxicity and related modes of toxic action was built by studying the establishment of toxicodynamic profiles, and behavioral and dose-response interpretation of 96-h LC_{50} tests (Russom et al., 1997). Good correlation between fish toxicity and toxic potency measures in *T. pyriformis* was also found in few studies (Bearden & Schultz, 1998; Schultz, 1997). McFarland (1970) proposed quite simple toxicokinetic- and toxicodynamic-based approach in which chemical toxicity was expressed as a combination of uptake into or through biological membranes and the interaction of the toxicant with the site of action to predict toxicity. Different approaches in modeling QSTR were also reported in the literature (Chakraborty et al., 2013; Cronin & Schultz, 1996; Karabunarliev et al., 1996; Veith & Mekenyan, 1993). QSTR could also be applied to measure toxicity of disinfection by-products (DBPs) generated during disinfection of water (Moudgal et al., 2000).

Cronin et al. (1998) developed QSTR model by using LUMO energy and maximum acceptor superdelocalizability as descriptors to predict toxicity of various alkyl- and halogen-substituted nitro- and dinitrobenzenes towards *T. Pyriformis*. Mishra et al. (2014) employed various quantum-chemical descriptors like HOMO energy, electronegativity, electron affinity (EA), ionization potential (IP), total energy and log P to predict the toxicity of nitrobenzene derivatives towards *T. Pyriformis*. The topological descriptors were also found to be effective to describe the toxicity of nitrobenzene derivatives towards the same (Tripathi & Mishra, 2014). Further based on several energy descriptors, heat of formation, atomic charge, dipole moment and polarizability values, Bao et al. (2012) predicted the toxicity of several nitrobenzene derivatives towards the *Scenedesmus obliguus*. Very recently, Tawari et al. (2014) attempted to build QSTR model to describe the mutagenicity of nitroaromatic compounds. Pandith and Islam (2013) further built QSTR models by using log P, electrophilicity (ω) and LUMO energy to predict the acute toxicity of aliphatic compounds towards *V. fischeri*. HOMO energy, molecular weight, LUMO energy, chemical potential, zero point energy, and hardness were also used to model QSTR in order to predict 50% effective inhibition concentration of various substituted benzene derivatives towards the *algae Scenedesmus obliquus* (Haghdadi & Fatemi, 2010). Quite high correlation ($r = 0.91$) revealed that such properties can be properly described by using such molecular indices. Eroglu et al. (2007) used quantum-chemical descriptors like translational entropy, principal moment of inertia, the highest and the lowest vibrational wave numbers to predict the acute toxicity of several organic compounds towards *Fathead Minnow*. Further, toxicity of organophosphorus compounds was assessed using the regression models mady by different quantum-chemical and topological descriptors (Senior et al., 2011). Wiener index, Harary index, modified Randic index, Kupchik modified connectivity index (Todeschini & Consonni, 2000), electronegative Randic index, heteroatom corrected extended connectivity Randic index

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

as different topological descriptors and HOMO-LUMO gap, maximum positive charge, total negative charge, charge polarization, local dipole index, electronic topological index, etc. as quantum-chemical descriptors were used for this purpose. Attempt to predict the lethal toxicity of organophosphorous pesticides occurred in the tadpoles was also reported in the literature (Yan et al., 2008). Antifungal and insecticidal activities of perfluorophenyl antimony(III) and antimony(V) chlorides were predicted by several energy and CDFT based descriptors by Singhal et al. (2014). Regression model based on total energy and LUMO energy was also found effective in describing the toxicity of several substituted benzenes against *freshwater photobacteria* (Mo et al., 2011). The hardness, electronegativity and total energy descriptors were found to be useful descriptors for QSTR modeling on a series of aliphatic alcohol derivatives (Kumar et al., 2011). Trohalaki and Pachter (2003) showed that the use of reciprocal of LUMO energy could improve the correlation in predicting the toxicity of halogenated aliphatic hydrocarbons than that of LUMO energy. The acute fish toxicity of organic solvents was also described by using the octanol-water partition coefficient, LUMO energy, dielectric constant and surface tension as descriptors (Levet et al., 2013). Gironés and Carbó-Dorca (2006) reported the molecular quantum similarity measures approach in develping QSTR.

3. CONCEPTUAL DENSITY FUNCTIONAL THEORY

The conceptual density functional theory (CDFT) (Chattaraj, 2009; Chattaraj & Giri, 2009; Chakraborty et al., 2010; Geerlings et al., 2003) has become a powerful theoretical tool to shed light into the chemical reactivity of a molecule. Within the framework of the CDFT, several response functions were introduced in order to describe chemical reactivity, which are also called as reactivity descriptors. CDFT along with its different global reactivity descriptors viz., hardness (η) (Parr & Pearson, 1983; Pearson, 1997), electronegativity (χ) (Chattaraj, 1992; Parr et al., 1978) and electrophilicity (ω) (Parr et al., 1999; Chattaraj et al., 2006, 2007, 2011) and different local reactivity descriptors like Fukui functions (Parr & Yang, 1984) $f(\mathbf{r})$ and their condensed-to-atom variants (Yang & Mortier, 1986) (f_k), local softness ($s(\mathbf{r})$) (Wang & Parr, 1985; Berkowitz & Parr, 1988), local hardness ($\eta(\mathbf{r})$) (Wang & Parr, 1985; Ghosh & Berkowitz, 1985) and philicity ($\omega(\mathbf{r})$) (Chattaraj et al., 2003; Roy et al., 2006a) predict the reactivity behavior of a molecule as a whole and / or at a particular site.

The concept of electronegativity (χ) was first introduced by Pauling (1932, 1960), which represents the "power" of an atom in a molecule to attract the bonded electrons towards itself. Later, Mulliken (1934, 1935) described χ as the average of the first ionization potential (*IP*) and the electron affinity (*EA*), two experimentally obtainable parameters of a molecule. He showed that χ could be determined by an arithmetic mean of the corresponding *IP* and *EA* values as given in Equation 1.

$$\chi = \frac{IP + EA}{2}. \tag{1}$$

The chemical potential (μ) (Parr & Yang, 1989; Hohenberg & Kohn, 1964) of an atom/molecule may be quantitatively assessed by the theory of statistical ensembles. Suppose, an atom or a molecule is a part of a grand canonical ensemble in which energy (E) and the number of electrons (N) are two independent parameters and continuous functions as well, then μ of the ensemble can be expressed as follows:

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

$$\mu = \frac{\partial E}{\partial N}\text{, at constant entropy} \tag{2}$$

Further, Gyftopoulos and Hatsopoulos (1968) put forward a definition of χ of a system where the electronic chemical potential (μ) was considered as the negative of χ. Now, following Mulliken's (1934, 1935) prescription for χ, the Mulliken chemical potential can also be expressed as a finite difference approximation of the derivative of E with respect to N as:

$$\mu_{Mulliken} = -\chi_{Mulliken} = -\frac{IP + EA}{2} = \left[\frac{\partial E(N)}{\partial N}\right]_{N=N_0} . \tag{3}$$

Pearson (1963, 1966, 1968a, 1968b, 1987) introduced the concept of "hardness" or "softness" in chemistry during the study of Lewis acid-base reactions

$$\text{A} + :\text{B} \rightarrow \text{A}:\text{B}$$

where A, an acid, acts as an electron acceptor whereas B, a base is the electron-pair donor. Global hardness (η) implies the extent of compactness of the electron cloud encompassing the nucleus/nuclei of an atomic/molecular system whereas global softness (S) is the reciprocal of η and hence represents the extent to which the electronic environment surrounding the nucleus/nuclei of an atomic/molecular species tends to loosen itself. Parr and Pearson (1983) expressed the qualitative concept of η quantitatively in the form of a mathematical descriptor. It is defined as the first derivative of μ or the second derivative of E as a function of N at a fixed external potential $v(\mathbf{r})$ as follows:

$$\eta = \left(\frac{\partial \mu}{\partial N}\right)_{v(\mathbf{r})} = \left(\frac{\partial^2 E}{\partial N^2}\right)_{v(\mathbf{r})} . \tag{4}$$

The convexity in the E vs. N curve signifies that the value of η would be always positive and from the finite difference approach the corresponding curvature equals to difference between IP and EA, which actually represents hardness. Hence

$$\eta = \left(\frac{\partial \mu}{\partial N}\right)_{v(\mathbf{r})} = \left(\frac{\partial^2 E}{\partial N^2}\right)_{v(\mathbf{r})} = IP - EA\,. \tag{5}$$

Therefore, η of a given system describes the extent of its reluctance towards the electron shift to another system in a general charge-transfer phenomenon in between a donor-acceptor couple.

Global softness (S) (Pearson, 1987), a reciprocal of the global hardness (η) of a system, is expressed as:

$$S = \frac{1}{\eta} = \left(\frac{\partial N}{\partial \mu}\right)_{v(\mathbf{r})} . \tag{6}$$

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Electrophilicity (ω) (Parr et al., 1999; Chattaraj et al., 2006, 2007, 2011) and its recently reported mathematical refinement, net electrophilicity ($\Delta\omega^{\pm}$) (Chattaraj et al., 2009) are also important descriptors from the perspective of constructing useful SAR based models

In chemistry electrophiles represent those species, which have a special affinity towards accepting electrons. In general, electrophiles are electron-deficient species, which prefer to combine with electron-rich systems, nucleophiles. In an experiment of the human immunodeficiency virus type 1 (HIV-1) nucleocapsid protein p7 (NCp7) with various electrophiles, Maynard et al. (1998) found that the decay rates of fluorescence varied almost linearly with the ratio of the square of χ and η, χ^2/η. This ratio could further be rationalized as the ability of an electrophile to attract electrons from a nucleophile in forming a covalent bond. This observation of Maynard et al. (1998) led Parr et al. (1999) to describe electron-attracting power of a system in a more precise way.

Consequently, a new reactivity descriptor namely electrophilicity index (ω) was introduced in chemistry. Different from the approach of Maynard et al., (1998) Parr's postulate of ω was based on a thermodynamic background and thus ω becomes a measure of the favorable change in energy upon saturation of a system with electrons. ω as prescribed by Parr et al. (1999) is:

$$\omega = \frac{\mu^2}{2\eta} = \frac{\chi^2}{2\eta}. \tag{7}$$

Eqution 7 resembles $W = V^2/R$, where W, V and R are electrical power, voltage and electrical resistance, respectively. Hence, ω describes the electrophilic power of a system. Few reviews summarizing the different application of ω in understanding the chemical reactivity of different systems are also available in literature (Chattaraj et al., 2006, 2007, 2011).

In order to correlate the associated energy changes between the interacting acceptors and donors in a charge-transfer process, based on the second-order Taylor series energy expansion formula as a function of the number of electrons (N) in the intervals between N - 1 and N, and N and N + 1, Gazquez et al. (2007) showed that the electrodonating (ω^-) and the electroaccepting (ω^+) powers might be defined as:

$$\omega^- = \frac{(\mu^-)^2}{2\eta^-} \tag{8a}$$

$$\omega^+ = \frac{(\mu^+)^2}{2\eta^+} \tag{8b}$$

where μ^- and μ^+ are the chemical potential for electron donation and electron acceptance, respectively, whereas η^- and η^+ are the hardness for electron donation and electron acceptance, respectively. It was also proved that μ^- and μ^+ could be equalized to μ and η^- and η^+ could be equalized to η. Therefore, ω^- and ω^+ become equivalent with ω and thereafter ω, ω^- or ω^+ may be represented in terms of μ and η as $\omega^- = \omega^+ = \omega = \mu^2/2\eta$. Further, Gazquez et al. (2007) proposed two different expressions for ω^- or ω^+ depending on two different approaches; one following the original formula as above and can be expressed as:

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

$$\omega^{+} = \frac{EA^{2}}{2(IP - EA)} \tag{9}$$

$$\omega^{-} = \frac{IP^{2}}{2(IP - EA)} \tag{10}$$

and other by exploiting an alternative expression for energy as (Gazquez et al., 2007):

$$\omega^{+} = \frac{(IP + 3EA)^{2}}{16(IP - EA)} \tag{11}$$

$$\omega^{-} = \frac{(3IP + EA)^{2}}{16(IP - EA)}. \tag{12}$$

It was also reported that the alternative expression of ω^{-} or ω^{+} provided better correlations than the original one (Gazquez et al., 2007).

Chattaraj et al. (2009) recently put forward the concept of net electrophilicity ($\Delta\omega^{\pm}$) to describe the resultant electron-accepting power of a molecule arising from the combined attractive and repulsive effects due to the presence of both electrons and nuclei within a molecule. This dual descriptor ($\Delta\omega^{\pm}$) assesses the electrophilicity of a molecule relative to its own nucleophilicity, therefore is more meaningful descriptor in representing the "electrophilic power" of a system. Thus, net electrophilicity ($\Delta\omega^{\pm}$) in terms of ω^{+} and ω^{-} can be expressed as:

$$\Delta\omega^{\pm} = \left\{\omega^{+} - \left(-\omega^{-}\right)\right\} = \left(\omega^{+} + \omega^{-}\right) \tag{13}$$

or,

$$\Delta\omega^{\pm} = \left\{\omega^{+} - \left(\frac{1}{\omega^{-}}\right)\right\}.$$

The polarizability (α) of an atom or a molecule is the lowest order response of its electron cloud with respect to an external weak electric field (Dalgarno, 1962; van Vleck, 1932). The static dipole polarizability (α) can be defined as the second derivative of the total electronic energy (E) with respect to the external homogeneous electric field as:

$$\alpha_{xy} = -\left(\frac{\partial^{2} E}{\partial F_{x} \partial F_{y}}\right)_{F=0} \tag{14}$$

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

where F_x and F_y are the electric field components along x and y direction. Qualitatively, polarizability (α) varies inversely with η (Ghanty & Ghosh, 1993, 1996; Fuentealba & Reyes, 1993; Pal & Chandra, 1995: Pearson, 1986; Politzer, 1987) and is proportional to S. A linear correlation was found between α and the third power of S (Chattaraj & Poddar, 1998; Simon-Manso & Fuentealba, 1998; Vela & Gázquez, 1990).

The magnetizability (ξ) (Griffith, 1999) for a molecule is a measure of the linear response of its electron cloud with respect to an externally applied magnetic field and is expressed as:

$$\left|\xi = -\left(\frac{\partial^2 \varepsilon(B)}{\partial B^2}\right)\right|_{B=0} \tag{15}$$

where B is the external magnetic field.

A softer molecule is also found to be more polarizable and hence more magnetizable (Chattaraj et al., 2007).

In addition to the various CDFT based global reactivity descriptors, the local reactivity descriptors like Fukui function ($f(\mathbf{r})$), philicity ($\omega(\mathbf{r})$), local softness ($s(\mathbf{r})$), local hardness ($\eta(\mathbf{r})$) and electron density ($\rho(\mathbf{r})$) are also important in providing valuable information regarding the reactivity pattern at a local atomic site in a molecule.

Information about $\rho(\mathbf{r})$ is important to get insights into site-selectivity for a molecule and can be determined in terms of the first order variation in energy (E) as a function of the external potential ($v(\mathbf{r})$) (Parr & Yang, 1989, 1995).

$$\rho(\mathbf{r}) = \left(\frac{\delta E}{\delta v(\mathbf{r})}\right)_N . \tag{16}$$

The Fukui function ($f(\mathbf{r})$) (Ayers & Levy, 2004; Fukui et al., 1952, 1954; Fukui, 1975, 1982; Parr & Yang, 1984; Yang et al., 1984) is very useful to evaluate quanlitatively the extent of local response at a particular atomic site of a molecule. It is defined as the differential change in $\rho(\mathbf{r})$ due to an infinitesimal change in the number of electrons (N).

$$f(\mathbf{r}) = \left(\frac{\partial \rho(\mathbf{r})}{\partial N}\right)_{v(\mathbf{r})} = \left(\frac{\delta \mu}{\delta v(\mathbf{r})}\right)_N . \tag{17}$$

Now, due to the discontinuity in the derivative in eq. 17, for integral values of N, by applying the finite difference and frozen core approximations three different types of Fukui functions can be defined as follows: (Ayers & Levy, 2004; Parr & Yang, 1984)

For nucleophilic attack:

$$f^+(\mathbf{r}) = \left(\frac{\partial \rho(\mathbf{r})}{\partial N}\right)^+_{v(\mathbf{r})} \approx \rho_{N+1}(\mathbf{r}) - \rho_N(\mathbf{r}) \approx \rho_{LUMO}(\mathbf{r}) . \tag{18a}$$

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

For electrophilic attack:

$$f^{-}(\mathbf{r}) = \left(\frac{\partial \rho(\mathbf{r})}{\partial N}\right)^{-}_{v(\mathbf{r})} \approx \rho_{N}(\mathbf{r}) - \rho_{N-1}(\mathbf{r}) \approx \rho_{HOMO}(\mathbf{r}). \tag{18b}$$

For radical attack:

$$f^{0}(\mathbf{r}) = \frac{1}{2}\left[f^{+}(\mathbf{r}) + f^{-}(\mathbf{r})\right]. \tag{18c}$$

Yang and Mortier (1986) further proposed a coarse-grained atom by atom representation of the Fukui function, called condensed-to-atom Fukui function. Based on a finite-difference approach, they can be represented as:

$$f_k^{+} = q_k(N+1) - q_k(N) \text{ for nucleophilic attack} \tag{19a}$$

$$f_k^{-} = q_k(N) - q_k(N-1) \text{ for electrophilic attack} \tag{19b}$$

$$f_k^{0} = [q_k(N+1) - q_k(N-1)] / 2 \text{ for radical attack} \tag{19c}$$

where q_k is the electron population at a particular k^{th} atomic site in a molecule.

Therefore, condensed-to-atom Fukui function can be easily computed through the data obtained from the population analysis.

Mathematically, global softness (*S*) can be expressed as an integral of the individual local softness (*s(***r***)*). A normalized condition between *S* and *s(***r***)* can be written as:

$$S = \int s(\mathbf{r}) dr \tag{20}$$

*s(***r***)* is connected to *f(***r***)* through a chain rule as:

$$s(\mathbf{r}) = \left(\frac{\partial \rho(\mathbf{r})}{\partial \mu}\right)_{v(\mathbf{r})} = \left(\frac{\partial \rho(\mathbf{r})}{\partial N}\right)_{v(\mathbf{r})} \left(\frac{\partial N}{\partial \mu}\right)_{v(\mathbf{r})} = f(\mathbf{r}).S \tag{21}$$

*f(***r***)* acts mainly as an intra-molecular reactivity descriptor by describing the reactivity of the atomic sites in a particular molecule whereas unlike *f(***r***)*, *s(***r***)* acts as an inter-molecular reactivity descriptor by comparing and correlating the tendency of interaction between a pair of neighboring molecules during chemical response.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Local hardness ($\eta(\mathbf{r})$) (Ghosh & Berkowitz, 1985; Yang & Parr, 1985) quantified the unwillingness of a particular atomic site in a molecule towards chemical attack. However, $\eta(\mathbf{r})$ cannot be normalized to η as was done for $s(\mathbf{r})$ to S. $\eta(\mathbf{r})$ can be expressed as:

$$\eta(\mathbf{r}) = \frac{1}{N}\int \frac{\delta^2 F[\rho(\mathbf{r})]}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')}\rho(\mathbf{r}')d\mathbf{r}' \tag{22}$$

where F[ρ(**r**)] is the Hohenberg-Kohn universal functional. $\eta(\mathbf{r})$ can also be expressed as:

$$\eta(\mathbf{r}) = \left(\frac{\delta\mu}{\delta\rho(\mathbf{r})}\right)_{v(\mathbf{r})}. \tag{23}$$

However, due to the dependence of $\rho(\mathbf{r})$ on $v(\mathbf{r})$, the above expression is ambiguous and hence cannot be considered as a basic definition for $\eta(\mathbf{r})$. Unlike η and S, $\eta(\mathbf{r})$ and $s(\mathbf{r})$ do not have a reciprocal relationship. They can be linked as:

$$\int \eta(\mathbf{r})s(\mathbf{r})dr = 1 \tag{24}$$

Further, the global hardness can be obtained from $\eta(\mathbf{r})$ and $f(\mathbf{r})$ as: (Chattaraj et al., 2007)

$$\eta = \int \eta(\mathbf{r})f(\mathbf{r})dr \tag{25}$$

Chattaraj et al. (2003) introduced the concept of philicity ($\omega^{\alpha}(\mathbf{r})$), which represents the aptitude of a local atomic site in a molecule towards towards reactivity (electrophilicity or nucleophilicity).

The relation between the local electrophilicity ($\omega(\mathbf{r})$) and the global electrophilicity (ω) can be expressed as:

$$\omega = \omega\int f(\mathbf{r})dr$$

$$\text{or, } \omega = \int \omega f(\mathbf{r})dr = \int \omega(\mathbf{r})dr \text{ thus, } \omega(\mathbf{r}) = \omega f(\mathbf{r}). \tag{26}$$

Therefore, from the expression it is obvious that $\omega(\mathbf{r})$ includes information about both ω and $f(\mathbf{r})$ but without the knowledge of ω, $f(\mathbf{r})$ alone cannot give any clue about $\omega(\mathbf{r})$. One can apply this concept of $\omega(\mathbf{r})$ in several types of chemical reaction. Therefore, it will be better to represent philicity as $\omega^{\alpha}(\mathbf{r})$ in stead of $\omega(\mathbf{r})$. α has signs of +, – and 0 describing nucleophilic, electrophilic and radical attacks, respectively. Therefore:

$$\begin{aligned} &\omega^{\alpha}\left(\mathbf{r}\right) = \omega.f^{\alpha}\left(\mathbf{r}\right) \\ &and, \quad \omega_k^{\alpha} = \omega.f_k^{\alpha} \end{aligned} \tag{27}$$

where ω_k^{α} is the condensed-to-atom local philicity variant for the k^{th} atomic site in a molecule.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

The concept of group philicity (ω_g^{α}) (Parthasarathi et al., 2004a) was also put forward in order to describe the reactivity of a group or an atomic assembly. It is generally expressed as a summation of the individual ω_k^{α} over all the related atoms. Therefore:

$$\omega_g^{\alpha} = \sum_{k=1}^{n} \omega_k^{\alpha} \tag{28}$$

where n describes the number of atoms present in the reacting group and $\alpha = +, -$ and 0 represents nucleophilic, electrophilic and radical attacks, respectively.

The efficacy of various CDFT based global and local reactivity descriptors in predicting chemical reactivity are guided by the various electronic structure principles like the principle of maximum hardness (PMH) (Ayers & Parr, 2000; Chattaraj et al., 1995, 2000; Chattaraj, 1996; Pan et al., 2013; Pan & Chattaraj, 2013; Parr & Chattaraj, 1991; Parr & Zhou, 1993; Pearson, 1987, 1993), minimum electrophilicity principle (MEP) (Chamorro et al., 2000; Pan et al., 2013; Parthasarathi, 2005), minimum polarizability principle (MPP) (Chattaraj & Sengupta, 1996, 1997; Chattaraj et al., 1999), minimum magnetizability principle (MMP) (Tanwar et al., 2006).

With the structural changes and/or substituent effects in a system, the variation of related conceptual DFT based reactivity descriptors led light into the quantitative correlation between chemical reactivity and toxicity. They indeed become important tools in building effective regression models for QSAR/QSPR/QSTR analyses.

4. ELECTROPHILICITY INDEX, NET ELECTROPHILICITY INDEX, AND THEIR LOCAL VARIANTS IN PREDICTING TOXICITY

Electrophilicity index (ω), net electrophilicity index ($\Delta\omega^{\pm}$) and the local variants of ω were found to be very effective tools in predicting toxicity. Some examples are given below.

4.1 Prediction of Toxicity of Various Alkali, Alkaline-Earth, Transition-Metal Ions and Arsenic Ions

Roy et al. (2009) attempted to predict the toxicity of arsenic (As^{3+} and As^{5+}) ions by choosing a training set composed of various alkali and alkaline-earth metal ions (Li^+, Na^+, K^+, Mg^{2+}, Ca^{2+}) and transition metal ions (Cr^{3+}, Mn^{2+}, Fe^{3+}, Co^{2+}, Ni^{2+}, Cu^{2+}, Zn^{2+}) along with their experimental toxicity (50% Lethal concentration, LC_{50}) in order to get single-parameter regression models. Two training sets composed of experimental toxicity (50% Lethal dose, LD_{50}) (Leermakers et al., 2006; Hughes, 2002) of two different sets of arsenic derivatives were also considered in order to predict their toxicity by employing several selected parameters. The corresponding atomic number (Z), energy (E), electrophilicity (ω) and the experimental as well as the calculated toxicity values expressed in terms of LC_{50} at the B3LYP/6-31+G(d) level are provided in Table 1. The calculated LC_{50} values in terms of Z and ω were computed as follows:

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 1. Atomic number (Z), energy (E, au), electrophilicity index (ω, eV), and experimental and calculated LC_{50} values for various selected alkali, alkaline-earth, and transition-metal ions along with arsenic ions at the B3LYP/6-31+G(d) level

Metal Ions	*Z*	*E*	*ω*	Expt. LC_{50}	Calc. LC_{50}	
					Z	*ω*
Training set						
Li^{+} [a]	3	−7.28459	22.033	0.215	–	0.283
Na^{+}	11	−162.08125	16.607	0.398	0.328	0.287
Mg^{2+}	12	−199.22743	37.645	0.250	0.310	0.272
K^{+} [a]	19	−599.72502	12.909	0.479	–	0.290
Ca^{2+}	20	−676.86695	28.515	0.173	0.165	0.278
Cr^{3+}	24	−1042.22188	495.511	0.019	0.093	−0.053
Mn^{2+}	25	−1149.83984	168.557	0.132	0.075	0.179
Fe^{3+}	26	−1261.27545	413.554	0.0003	0.056	0.005
Co^{2+}	27	−1381.56785	204.105	0.022	0.038	0.154
Ni^{2+}	28	−1507.03577	331.966	0.069	0.020	0.063
Cu^{2+} [b]	29	−1639.25180	100.906	0.002	0.002	–
Zn^{2+} [b]	30	−1778.15610	65.992	0.007	−0.016	–
Toxicity prediction of arsenic ions						
As^{3+}	33	−2231.61591	115.148	–	–	0.217
As^{5+}	33	−2227.48829	182.990	–	–	0.169

[a]Outliers in the Z model;[b]Outliers in the ω model.

(Reprinted with permission from Roy et al., 2009. Copyright © 2009, Springer Science+Business Media B.V.).

$$\text{Calc. } LC_{50} = -0.0181 \times Z + 0.5272; \tag{29}$$

$$R = 0.923,\ SD = 0.054,\ N = 10$$

$$\text{Calc. } LC_{50} = -7.1 \times 10{-}04 \times \omega + 0.2987; \tag{30}$$

$$R = 0.793,\ SD = 4.337,\ N = 10.$$

In both *Z* and *ω* models, two ions were found to be outliers, being Li^{+} and K^{+} in case of *Z*- model and Cu^{2+} and Zn^{2+} in case of *ω*- model. However, the *Z*- model was found to be better in prediction of toxicity of ions than that of ω- model but being the same *Z* value, Z- model could not differentiate the toxicity of As^{3+} and As^{5+}. On the other hand, *ω*- model can predict the toxicity of As^{3+} and As^{5+} correctly, As^{3+} ($LC_{50} = 0.217$) being more toxic than As^{5+} ($LC_{50} = 0.169$).

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Further, efficacy of different global descriptors like number of nonhydrogenic atoms (N_{NH}) and ω and different local descriptors like philicity (ω^{+}_{As}) and atomic charges (Q_{As}) at As site were tested in building QSTR taking two sets of As-derivatives, whose experimental toxicity were known in terms of LD_{50} obtained from two different biological assays (Leermakers et al., 2006; Hughes, 2002).

In both sets, seven As-derivatives were included (Set-I members were Arsenite (As(III)), Arsenobetaine (AsB), Arsenocholine (AsC), Monomethyl arsonic acid (MMA), Dimethyl arsinic acid (DMA), Trimethyl arsine oxide (TMAO), Tetramethylarsonium ion (TeMA) and Set-II members were Arsenate [As(V)], Arsenite [As(III)], Arsenobetaine (AsB), Arsenic trioxide (As_2O_3), Monomethyl arsonic acid (MMA), Dimethyl arsinic acid (DMA), Trimethyl arsine oxide (TMAO)). For each individual descriptor, one-parameter regression analyses were performed. It was found that N_{NH} has reasonably good predictive power towards the toxicity of As-derivatives in both the sets except the case of TMAO molecule.

ω showed better prediction in set-II than in set-I. On the other hand, ω^{+}_{As} showed better prediction in set-I than in set-II and Q_{As} could not predict the toxicity of As(III) in both the sets as well as As_2O_3 in set-II. Nevertheless, none of these descriptors could predict the toxicity of all seven molecules in both the sets. Then possible combinations of these four descriptors were tested towards the prediction of toxicity. Two-parameter regressions were not found that much suitable in predicting toxicity of all the As-derivatives; however, three-parameter regressions composed of one global (N_{NH} or ω) and two local parameters (ω^{+}_{As} and Q_{As}) were capable to represent the toxicity of all the derivatives reasonably well (see Figure 1).

The corresponding regression models could be represented as follows:

Set-I:

$$\text{Calc.LD}_{50} = 1.983 \times N_{\text{NH}} - 5.360 \times \omega^{+}_{\text{As}} - 35.926 \times Q_{\text{As}} + 15.402 \quad (31)$$

$R = 0.866$, $SD = 2.516$, $N = 7$

$$\text{Calc.LD}_{50} = 1.805 \times \omega - 16.312 \times \omega^{+}_{\text{As}} - 85.353 \times Q_{\text{As}} + 51.415 \quad (32)$$

$R = 0.898$, $SD = 2.216$, $N = 7$

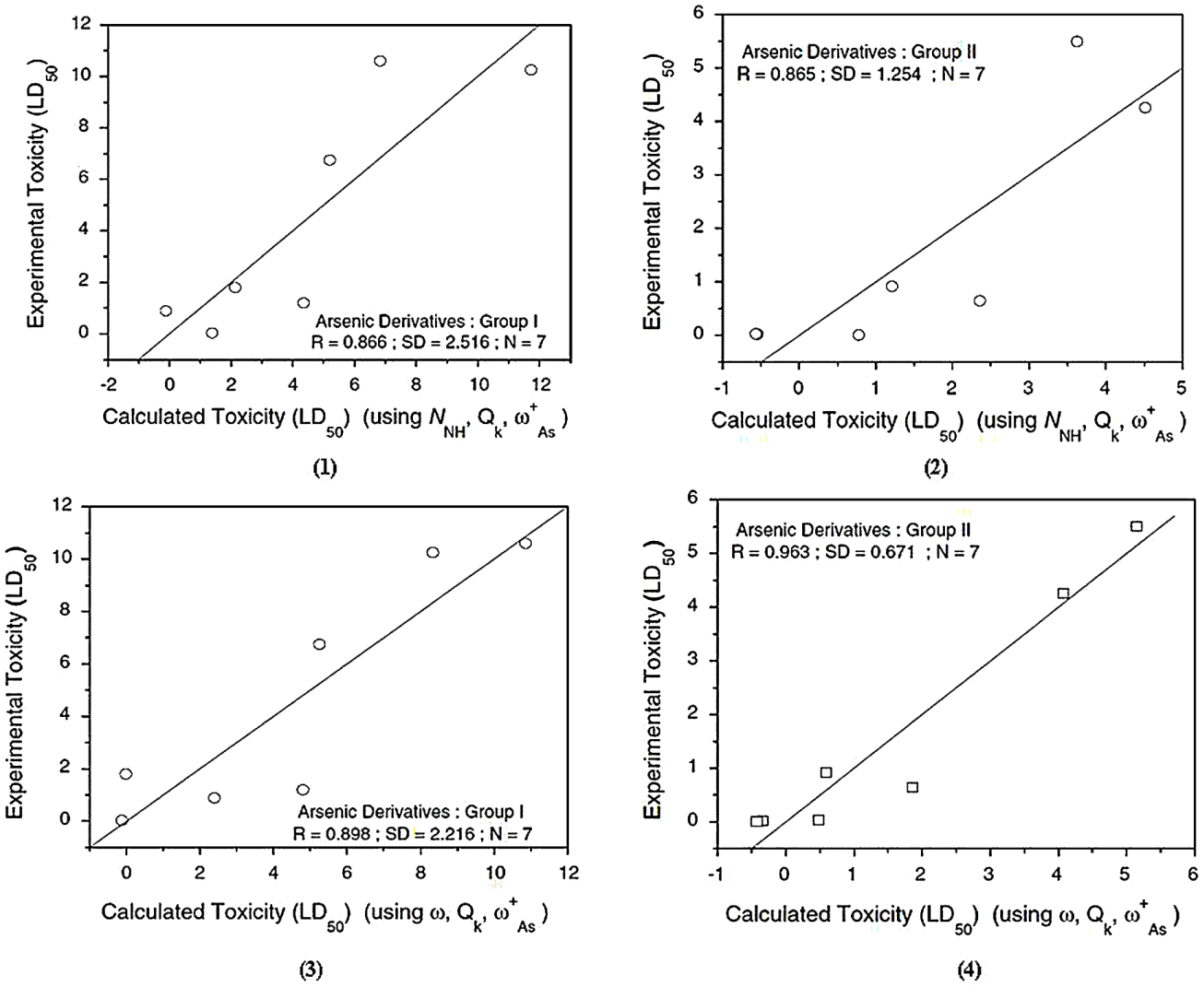
Set-II:

$$\text{Calc.LD}_{50} = 0.484 \times N_{\text{NH}} - 3.232 \times \omega^{+}_{\text{As}} - 16.968 \times Q_{\text{As}} + 10.065 \quad (33)$$

$R = 0.865$, $SD = 1.254$, $N = 7$

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

*Figure 1. Experimental versus calculated toxicity (LD50) values using three-parameter (N_{NH}, ω^{+}_{As}, and Q_{As}) regression model for the complete set of set-I (**1**) and II (**2**) As-derivatives and three-parameter (ω, ω^{+}_{As}, and Q_{As}) regression model for the complete set of set-I (**3**) and II (**4**) As-derivatives*
(Reprinted with permission from Roy et al., 2009. Copyright © 2009, Springer Science+Business Media B.V.).



$$\text{Calc.LD}_{50} = 4.724 \times \omega - 10.881 \times \omega^{+}_{\text{As}} - 54.889 \times Q_{\text{As}} + 23.380 \quad (34)$$

$R = 0.963$, $SD = 0.671$, $N = 7$

It was also found that the toxicity prediction power of the combination, ω, ω^{+}_{As} and Q_{As} is better than that of the combination of N_{NH}, ω^{+}_{As} and Q_{As}.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

4.2 Prediction of the Toxicity ($pIGC_{50}$) of 252 Aliphatic Compounds on *Tetrahymena Pyriformis*

Chattaraj et al. (2007) used global descriptors like the number of carbon atoms (N_C), number of nonhydrogenic atoms (N_{NH}) in a molecule, ω and local descriptors like ω^+_{max}, ω^-_{max}, q_k^{max} computed at the HF/6-311G(d) level in modeling effective QSTR of a large number of aliphatic compounds on *T. pyriformis*. Here ω^+_{max} and ω^-_{max} referred to the ω_k^{α} at an atomic site in which it was maximum. q_k^{max} implied atomic charge at the k^{th} site having maximum charge. By studying 252 aliphatic compounds, it was found that apart from the set of aliphatic amines, the atom number (N_C and N_{NH}) was an important descriptor whereas ω was also turned out as an effective descriptor in most of the cases. The predictability towards toxicity of two-parameter models based on N_C and ω was also found almost same with respect to the corresponding three-parameter models built by N_C, ω and any one descriptor from N_{NH}, ω^+_{max}, ω^-_{max}, q_k^{max}. Different regression equations adopted in one (N_C)- and two (N_C, ω)-parameter models are given Table 2. N_C-based single parameter model could predict toxicity fairly in most of the cases. In few cases like carboxylic acids, halogenated acids, amino alcohols and amines, it did not work. However, except amines the predictability towards toxicity of the model was found to improve significantly in coupling N_C with ω. In cases of amines, q_k^{max} and ω^-_{max} were also found to be important in describing toxicity. With the observation that in many cases, similar results were obtained, N_C was termed as the crude alternative of log P. The plots of experimental $pIGC_{50}$ versus calculated $pIGC_{50}$ values are displayed in Figure 2. Quite large R^2, R^2_{cv} and variance adjusted to degrees of freedom, R^2_{adj} values implied the effectiveness of these models in building QSTR.

4.3 Prediction of Toxicity of Several Polyaromatic Hydrocarbons (PAH)

The global and local electrophilicities computed at the B3LYP/6-31G(d) level were used to predict the toxicity of several polyaromatic hydrocarbons (PAH) (Roy et al., 2006c). The toxicity was represented in terms of pIC_{50}, which could be defined as the molar concentration of these chemicals required to displace 50% of radio-labeled tetrachlorodibenzo-p-dioxin (TCDD) from the arylhydrocarbon (Ah) receptor. As a training set, dependent and independent variables were considered as the experimental pIC_{50} values of the electron acceptor toxin like polychlorinated dibenzofurans (PCDF) and ω, respectively. Further, this model was also employed on another set of polychlorinated biphenyls (PCB). General structures of PCDFs, and PCBs are provided in Tables 3 and 4, respectively, along with the atom numbering, identitity number and substitution pattern. The fractional number of electrons (ΔN) shifted from the PCDFs, PCBs and amines (A) to the nucleic acid bases/DNA base pairs (B) was also used as a descriptor to describe toxicity. The positive values of ΔN for PCDFs and PCBs implied their electron accepting nature whereas the negative values for amines recognized them as electron donors. Reasonably good correlation was found to exist between ΔN (in between toxin and guanine) and ω having R = 0.991. The correlation between ΔN and exp. pIC_{50} was 0.881 whereas that obtained from the two-parameter regression based on ΔN and ω and exp. pIC_{50} was 0.892 (see Figure 3). Similarly, in cases of interactions in between PCDFs and GCWC (Guanine-Cystosine Watson-Crick) base-pairs, good correlations were observed. The calculated parameters (ω and ω^+_{max}) along with the experimental and predicted pIC_{50} values of the training set of PCDFs and PCBs are given in Table 5. In case of PCDFs, ω values gave a correlation R value of 0.891 whereas in case of PCBs, it was 0.834 (see Figure 4). The toxicity of aliphatic amines expressed by the 50% inhibitory growth concentration (IGC_{50}) towards ciliate fresh-water protozoa *Tet-*

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 2. Regression models for different groups of aliphatic compounds for estimating their toxicity towards Tetrahymena pyriformis

Molecules	Regression Equations	R	SD
Aliphatic electron acceptors			
Diols (N = 10)	$pIGC_{50} = 0{\cdot}4497 \times N_C - 4{\cdot}0855$	0·9683	0·2781
	$pIGC_{50} = 0{\cdot}8059 \times \log P - 1{\cdot}4688$	0·9892	0·1617
	$pIGC_{50} = -12{\cdot}4224 \times \omega + 0{\cdot}3554 \times N_C + 7{\cdot}6094$	0·9826	0·2070
	$pIGC_{50} = 10{\cdot}0678 \times \omega + 0{\cdot}9625 \times \log P - 10{\cdot}5043$	0·9934	0·1270
Halogenated alcohols (N = 11)	$pIGC_{50} = 0{\cdot}3271 \times N_C - 2{\cdot}0248$	0·8923	0·3852
	$pIGC_{50} = 0{\cdot}7783 \times \log P - 1{\cdot}3735$	0·9486	0·2561
	$pIGC_{50} = -6{\cdot}2863 \times \omega + 0{\cdot}1982 \times N_C + 4{\cdot}5793$	0·9424	0·2855
	$pIGC_{50} = -4{\cdot}0784 \times \omega + 0{\cdot}5772 \times \log P + 2{\cdot}7468$	0·9646	0·2169
Saturated alcohols (N = 32)	$pIGC_{50} = 0{\cdot}4144 \times N_C - 3{\cdot}0801$	0·9634	0·3456
	$pIGC_{50} = 0{\cdot}7745 \times \log P - 2{\cdot}0034$	0·9903	0·1777
	$pIGC_{50} = 0{\cdot}8927 \times \omega + 0{\cdot}4261 \times N_C - 3{\cdot}9484$	0·9636	0·3451
	$pIGC_{50} = 1{\cdot}6835 \times \omega + 0{\cdot}8138 \times \log P - 3{\cdot}5796$	0·9907	0·1739
Carboxylic acids (N = 28)	$pIGC_{50} = 0{\cdot}1116 \times N_C - 0{\cdot}9678$	0·6676	0·2917
	$pIGC_{50} = 0{\cdot}2857 \times \log P - 0{\cdot}7006$	0·9586	0·1070
	$pIGC_{50} = -5{\cdot}4426 \times \omega + 0{\cdot}0338 \times N_C + 4{\cdot}8562$	0·8801	0·1860
	$pIGC_{50} = -0{\cdot}3944 \times \omega + 0{\cdot}2715 \times \log P - 0{\cdot}2924$	0·9589	0·1066
Halogenated acids (N = 11)	$pIGC_{50} = 0{\cdot}2257 \times N_C - 1{\cdot}3481$	0·6564	0·3632
	$pIGC_{50} = 0{\cdot}4620 \times \log P - 0{\cdot}8744$	0·8107	0·2285
	$pIGC_{50} = 1{\cdot}8930 \times \omega + 0{\cdot}0976 \times N_C - 2{\cdot}2827$	0·9186	0·1903
	$pIGC_{50} = 1{\cdot}6012 \times \omega + 0{\cdot}2001 \times \log P - 1{\cdot}8388$	0·9169	0·1762
Mono esters (N = 31)	$pIGC_{50} = 0{\cdot}3645 \times N_C - 2{\cdot}7969$	0·9189	0·3710
	$pIGC_{50} = 0{\cdot}7599 \times \log P - 2{\cdot}0274$	0·9645	0·2396
	$pIGC_{50} = -10{\cdot}9131 \times \omega + 0{\cdot}2554 \times N_C + 8{\cdot}1384$	0·9352	0·3330
	$pIGC_{50} = -3{\cdot}0902 \times \omega + 0{\cdot}6960 \times \log P + 1{\cdot}0027$	0·9655	0·2365
Diesters (N = 20)	$pIGC_{50} = 0{\cdot}2861 \times N_C - 2{\cdot}884$	0·9299	0·3382
	$pIGC_{50} = 0{\cdot}6338 \times \log P - 1{\cdot}3322$	0·9539	0·2632
	$pIGC_{50} = -4{\cdot}8166 \times \omega + 0{\cdot}1999 \times N_C + 1{\cdot}1227$	0·9636	0·2460
	$pIGC_{50} = -4{\cdot}2407 \times \omega + 0{\cdot}4687 \times \log P + 1{\cdot}7763$	0·9790	0·1834
Aldehydes (N = 13)	$pIGC_{50} = 0{\cdot}2230 \times N_C - 1{\cdot}4027$	0·8980	0·2459
	$pIGC_{50} = 0{\cdot}4628 \times \log P - 0{\cdot}8864$	0·9227	0·1988
	$pIGC_{50} = -2{\cdot}5248 \times \omega + 0{\cdot}1228 \times N_C + 1{\cdot}3002$	0·9332	0·2008
	$pIGC_{50} = -2{\cdot}1731 \times \omega + 0{\cdot}2904 \times \log P + 1{\cdot}2280$	0·9496	0·1664
Ketones (N = 15)	$pIGC_{50} = 0{\cdot}4147 \times N_C - 3{\cdot}4470$	0·9850	0·2249
	$pIGC_{50} = 0{\cdot}7720 \times \log P - 2{\cdot}0314$	0·9872	0·2048
	$pIGC5_0 = -3{\cdot}2176 \times \omega + 0{\cdot}38989 \times N_C - 0{\cdot}6459$	0·9855	0·2211
	$pIGC_{50} = -1{\cdot}4487 \times \omega + 0{\cdot}7511 \times \log P - 0{\cdot}8080$	0·9873	0·2041

continued on following page

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 2. Continued

Molecules	Regression Equations	R	SD
Aliphatic electron donors			
Amino alcohols (N = 18)	$pIGC_{50} = 0{\cdot}2255 \times N_C - 2{\cdot}3296$	0·4711	0·4596
	$pIGC_{50} = 0{\cdot}3533 \times \log P - 1{\cdot}0529$	0·5829	0·2468
	$pIGC_{50} = 8{\cdot}9875 \times \omega + 0{\cdot}1464 \times N_C - 7{\cdot}4431$	0·9152	0·2100
	$pIGC_{50} = 8{\cdot}5520 \times \omega + 0{\cdot}2282 \times \log P - 6{\cdot}3481$	0·9377	0·1697
Acetylenic alcohols (N = 13)	$pIGC_{50} = 0{\cdot}3332 \times N_C - 2{\cdot}2125$	0·8942	0·4218
	$pIGC_{50} = 0{\cdot}5506 \times \log P - 0{\cdot}8071$	0·8842	0·3891
	$pIGC_{50} = 3{\cdot}7452 \times \omega + 0{\cdot}3707 \times N_C - 5{\cdot}0494$	0·9080	0·3947
	$pIGC_{50} = 0{\cdot}2523 \times \omega + 0{\cdot}5538 \times \log P - 0{\cdot}9858$	0·8843	0·3890
Unsaturated alcohols (N = 25)	$pIGC_{50} = 0{\cdot}4093 \times N_C - 3{\cdot}3473$	0·8580	0·3311
	$pIGC_{50} = 0{\cdot}7587 \times \log P - 1{\cdot}6861$	0·9315	0·2185
	$pIGC_{50} = -2{\cdot}2250 \times \omega + 0{\cdot}3641 \times N_C - 1{\cdot}7327$	0·9136	0·2622
	$pIGC_{50} = -0{\cdot}0271 \times \omega + 0{\cdot}7568 \times \log P - 1{\cdot}6679$	0·9315	0·2185
Amines (N = 25)	$pIGC_{50} = 0{\cdot}1136 \times N_C - 1{\cdot}3456$	0·2711	0·3965
	$pIGC_{50} = 0{\cdot}4609 \times \log P - 1{\cdot}1380$	0·8534	0·1833
	$pIGC_{50} = -0{\cdot}1700 \times \omega + 0{\cdot}1116 \times N_C - 1{\cdot}2295$	0·2723	0·3964
	$pIGC_{50} = 0{\cdot}9307 \times \omega + 0{\cdot}4792 \times \log P - 1{\cdot}7367$	0·8641	0·1792
	$pIGC_{50} = 0{\cdot}1162 \times q_k^{max} + 2{\cdot}1524 \times \omega^-_{max} + 0{\cdot}0669 \times N_C - 1{\cdot}8782$	0·8692	0·2037
	$pIGC_{50} = 0{\cdot}0681 \times q_k^{max} + 1{\cdot}3490 \times \omega^-_{max} + 0{\cdot}2802 \times \log P - 1{\cdot}4885$	0·9429	0·1293

(Reprinted with permission from Chattaraj et al., 2007.
Copyright © 2007, Indian Academy of Sciences).

rahymena pyriformis was also assessed. Due to the lack of global nucleophilicity, local nucleophilicity (ω^+_{max}) was applied as a descriptor for this set. This model was then further employed for the test set of amino alcohols. It was found that despite the occurrence of a very good correlation, there was some cross validation problem. Additionally, the use of the amines and amino alcohols together as both training and test sets provided fair correlation.

4.4 Structure-Toxicity Analysis on Chloroanilines

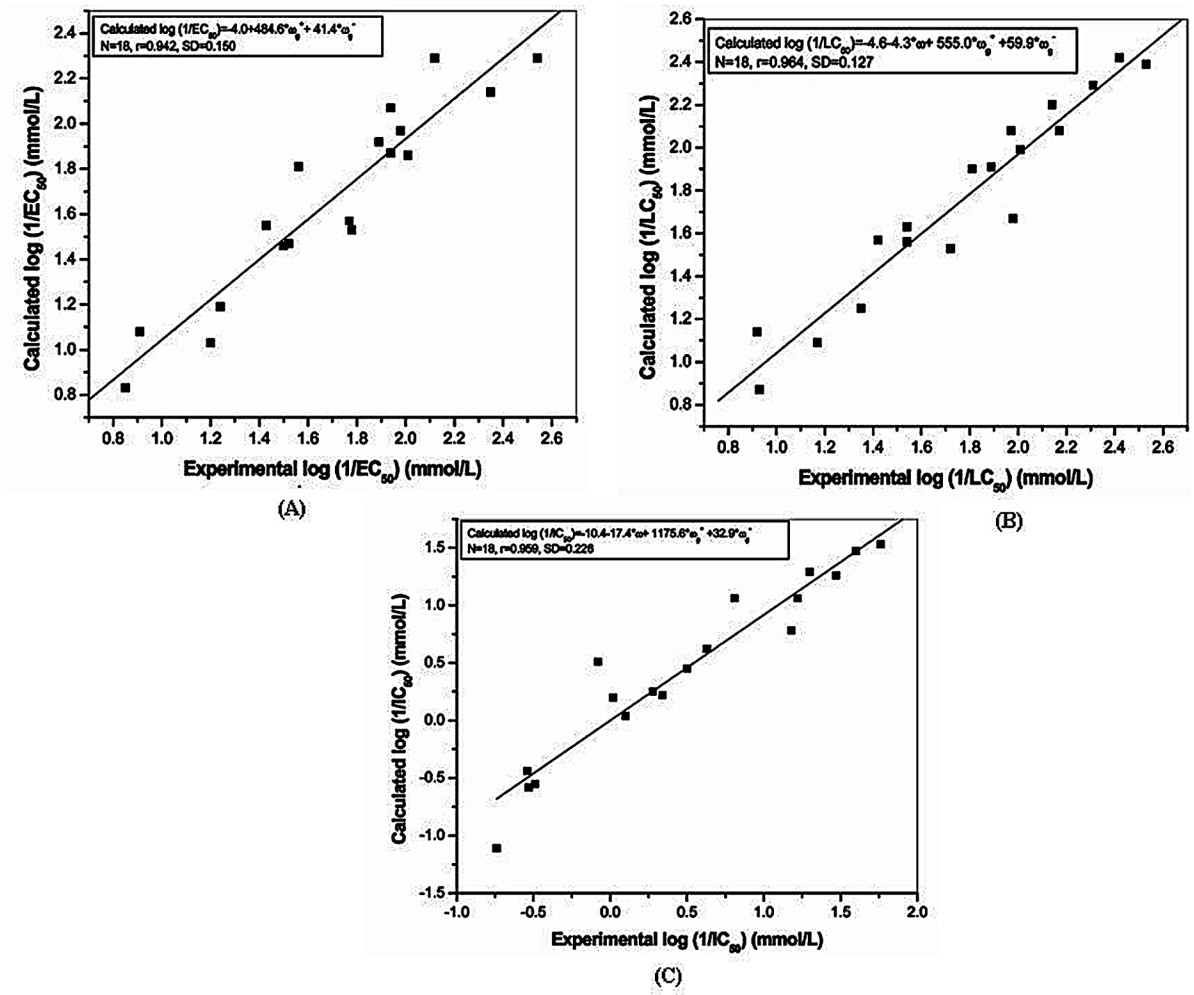
Padmanabhan et al. (2006a) further developed QSTR for aquatic toxicity ($\log(IGC_{50}^{-1})$) of a set of nine chloroanilines (CA) against *T. pyriformis* at the B3LYP/6-311++G(d,p) level as given in Table 6. The experimental and calculated $\log(IGC_{50}^{-1})$ are also given in Table 6. The regression equation considering experimental $\log(IGC_{50}^{-1})$ value (Schultz, 1999) as a dependent variable and ω as an independent variable was as follows:

$$\log(IGC_{50}^{-1}) = -4.414 + 56.035\ \omega \tag{35}$$

$N = 9$, $r^2 = 0.968$, SD = 0.144.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 2. Observed versus calculated $pIGC_{50}$ values using two-parameter (ω, N_C) regression models for the (a) Complete set of aliphatic electron acceptors and (b) complete set of aliphatic electron donors
(Reprinted with permission from Chattaraj et al., 2007. Copyright © 2007, Indian Academy of Sciences).



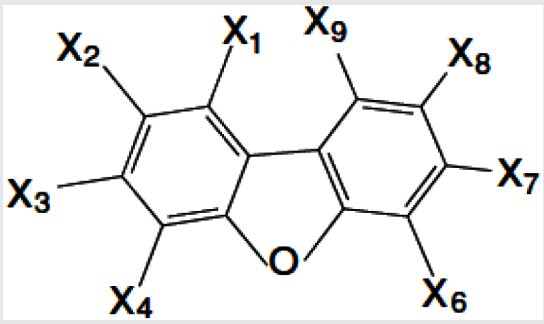
The plot between exp. and calc. $\log(IGC_{50}^{-1})$ (see Figure 5) showed that ω could describe the aquatic toxicity of these CA systems superbly having a correlation coefficient (r) of 0.984.

4.5 Prediction of Toxicity of Phenol, Nitrobenzene, and Benzonitrile-Derivatives

ω and its local variants, local electrophilicity index (ω_k^{α}) computed at the HF/6-31G(d) level were also employed in order to describe the toxicity of a large number (174) of aromatic compounds towards *T. pyriformis*, mainly belonging to the family of phenols, nitrobenzenes and benzonitriles (Roy et al., 2006b). At first, depending on their electron donating/accepting nature measured by comparing their χ values with those of the nuclei acid bases (adenine, thymine, guanine, cytosine and uracil) and DNA base pairs (GCWC and ATH), they were divided into two classes viz., 97 phenol derivatives having electron donating group and 77 nitrobenzenes and benzonitriles having electron accepting group. The regression models were used for the complete set of aromatic donors as

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 3. Polychlorinated dibenzofurans with identity number (ID) representing the substitution pattern

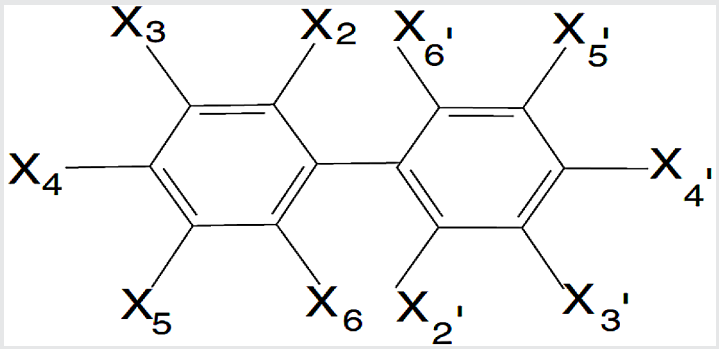


ID	X_1	X_2	X_3	X_4	X_6	X_7	X_8	X_9
1	H	Cl	H	H	H	H	H	H
2	H	H	H	Cl	H	H	H	H
3	H	Cl	H	H	Cl	H	H	H
4	H	Cl	H	H	H	H	Cl	H
5	Cl	H	Cl	H	Cl	H	H	H
6	Cl	H	Cl	H	H	H	Cl	H
7	H	Cl	Cl	Cl	H	H	H	H
8	H	Cl	Cl	H	H	H	Cl	H
9	H	Cl	H	H	Cl	Cl	H	H
10	H	Cl	Cl	Cl	Cl	H	H	H
11	H	Cl	Cl	Cl	H	H	Cl	H
12	H	Cl	Cl	H	H	Cl	Cl	H
13	Cl	Cl	H	Cl	Cl	Cl	H	H
14	Cl	Cl	Cl	Cl	H	H	Cl	H
15	Cl	Cl	Cl	H	H	Cl	Cl	H
16	H	Cl	Cl	Cl	H	Cl	Cl	H
17	Cl	Cl	Cl	Cl	H	Cl	Cl	H
18	H	Cl	Cl	Cl	Cl	Cl	Cl	H
19	H	Cl	Cl	H	Cl	H	Cl	H
20	Cl	Cl	Cl	H	Cl	H	H	H
21	Cl	Cl	Cl	H	H	Cl	H	H
22	Cl	H	Cl	Cl	H	Cl	Cl	H
23	H	Cl	Cl	Cl	H	Cl	H	Cl
24	Cl	Cl	Cl	H	H	Cl	H	Cl
25	Cl	Cl	Cl	H	H	Cl	H	H
26	Cl	H	Cl	Cl	H	Cl	Cl	H
27	H	Cl	Cl	Cl	H	Cl	H	Cl

(Reprinted with permission from Roy et al., 2006c. Copyright © 2006, Springer Science+Business Media, Inc.).

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

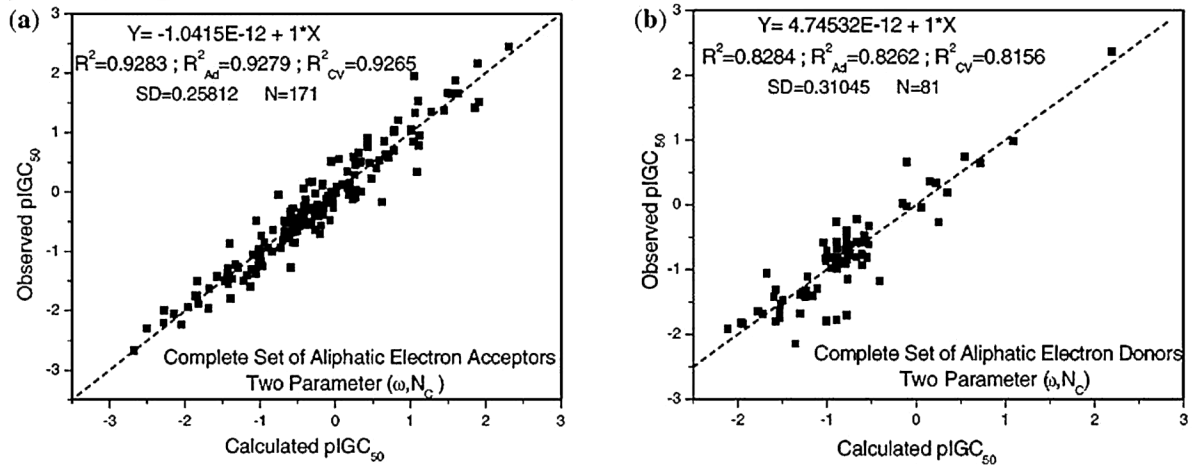
Table 4. Polychlorinated biphenyls with identity number (ID) representing the substitution pattern



ID	X_2	X_3	X_4	X_5	X_6	X_2'	X_3'	X_4'	X_5'	X_6'
28	Cl	H	H	H	H	H	Cl	Cl	Cl	H
29	Cl	Cl	Cl	H	H	H	Cl	Cl	H	H
30	Cl	H	Cl	Cl	H	H	Cl	Cl	H	H
31	Cl	Cl	Cl	Cl	H	H	Cl	Cl	H	H
32	Cl	H	Cl	H	H	Cl	H	Cl	H	H
33	Cl	H	Cl	H	Cl	H	Cl	Cl	Cl	H

(Reprinted with permission from Roy et al., 2006c. Copyright © 2006, Springer Science+Business Media, Inc.).

*Figure 3. Variation of charge transfer (ΔN) with (**A**) electrophilicity index (ω), (**B**) the experimental biological activity (pIC_{50}) during interaction of polychlorinated dibenzo furan with guanine and (**C**) Observed versus calculated pIC_{50} values of polychlorinated dibenzo furan using electrophilicity index (ω) and N with guanine (For **C**, the regression model is: $pIC_{50} = 4.033*\omega - 13.596*\Delta N - 5.954$)*
(Reprinted with permission from Roy et al., 2006c. Copyright © 2006, Springer Science+Business Media, Inc.).



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

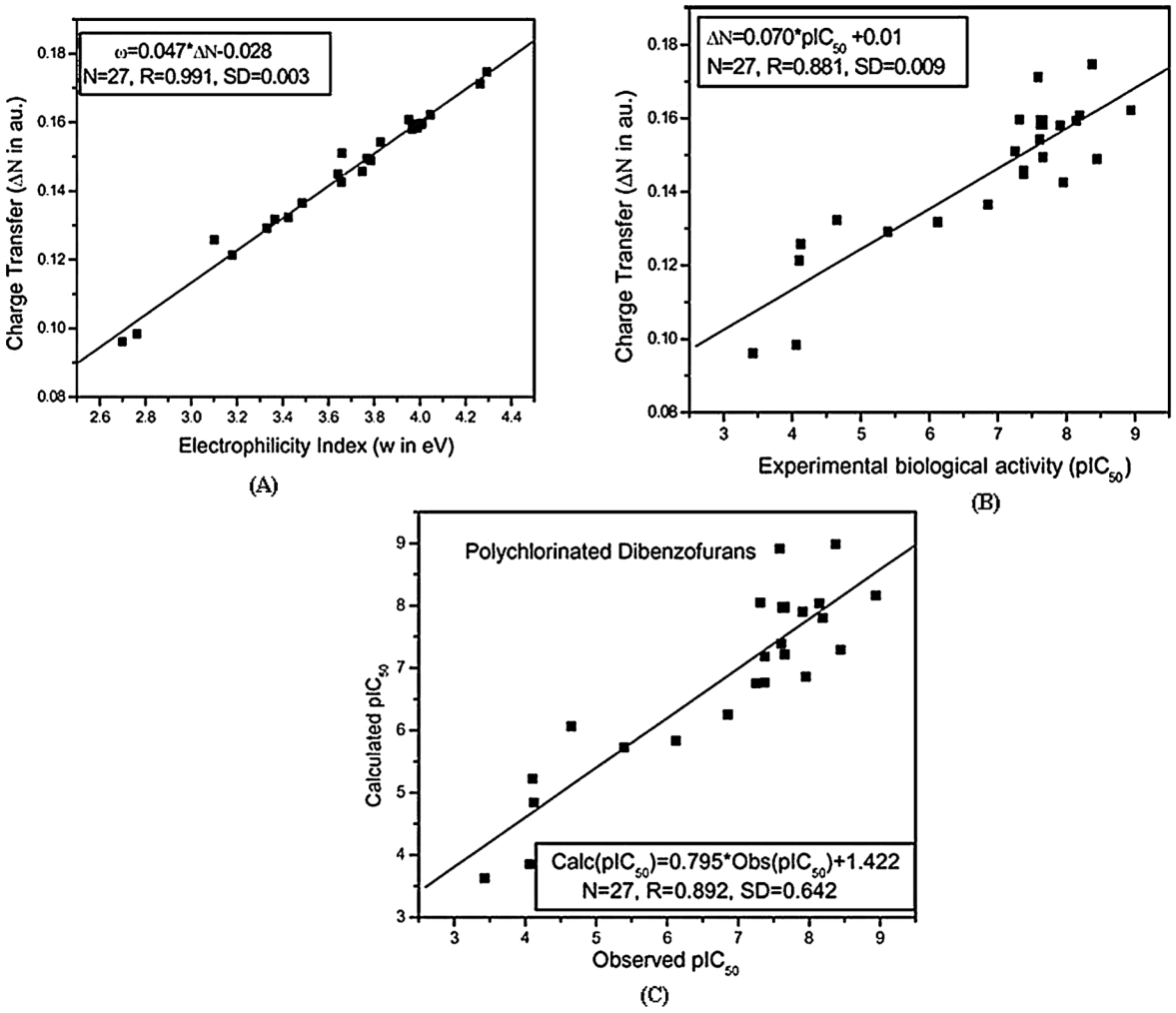
Table 5. Experimental and calculated biological activity (pIC_{50}) in gas phase for the training set of polychlorinated dibenzofurans and polychlorinated biphenyls

ID	ω (eV)	ω^{+}_{max} (eV)	Observed pIC_{50}	Calculated pIC_{50}
1	2.763	0.888	4.061	3.800
2	2.699	0.720	3.429	3.583
3	3.101	0.883	4.125	4.947
4	3.180	0.893	4.103	5.216
5	3.366	1.306	6.123	5.847
6	3.425	1.007	4.653	6.047
7	3.331	1.156	5.396	5.728
8	3.486	1.488	6.858	6.254
9	3.659	1.551	7.255	6.841
10	3.748	1.551	7.379	7.144
11	3.769	1.633	7.657	7.215
12	3.785	1.824	8.444	7.269
13	3.952	1.717	8.194	7.836
14	3.967	1.683	7.911	7.887
15	4.005	1.756	8.147	8.016
16	4.046	1.902	8.943	8.155
17	4.263	1.612	7.587	8.892
18	4.293	1.756	8.376	8.994
19	3.828	1.621	7.610	7.415
20	3.642	1.57	7.379	6.784
21	3.657	1.718	7.954	6.835
22	3.989	1.629	7.657	7.962
23	3.988	1.629	7.657	7.958
24	4.010	1.575	7.313	8.033
25	3.657	1.718	7.954	6.835
26	3.989	1.621	7.623	7.962
27	3.988	1.622	7.623	7.958
28	3.109	0.476	5.584	4.975
29	3.329	0.432	6.134	5.721
30	3.424	0.440	5.762	6.044
31	3.597	0.461	6.057	6.631
32	2.866	0.318	4.442	4.150
33	3.112	0.406	4.577	4.985

(Reprinted with permission from Roy et al., 2006c. Copyright © 2006, Springer Science+Business Media, Inc.).

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 4. (A) Calculated versus observed values of biological activity (pIC_{50}) for training set of PCDFs, (B) predicted versus observed values of biological activity (pIC_{50}) for test set of PCBs in gas phase
(Reprinted with permission from Roy et al., 2006c. Copyright © 2006, Springer Science+Business Media, Inc.).



$$\log(IGC_{50}^{-1}) = -0.0703\ \omega + 0.7706\ \omega^{-}_{O\max} - 6.10^{-05}E_{HF}\ 0.6179 \quad (36)$$

N=97, R=0.898, $R^2_{adj} = 0.804$, $R^2_{CV} = 0.795$, SD=0.290

and that for the complete set of aromatic acceptors as

$$\log(IGC_{50}^{-1}) = 0.6612\ \omega - 1.2673\ \omega^{+}_{C\max} - 4.10^{-05}E_{HF}\ 1.1752 \quad (37)$$

N=77, R=0.902, $R^2_{adj} = 0.811$, $R^2_{CV} = 0.800$, SD=0.359

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 6. Experimental and calculated toxicity (log(IGC_{50}^{-1})) values against T. pyriformis of the selected chloroanilines at the B3LYP/6-311++G(d,p) method

System	Log(IGC_{50}^{-1})		Residual[b]
	Observed[a]	Calculated	
2-CA	-0.17	-0.09	-0.08
3-CA	0.22	-0.02	0.24
4-CA	0.05	0.01	0.04
2,5-C2A	0.58	0.72	-0.14
3,4-C2A	0.56	0.67	-0.11
3,5-C2A	0.71	0.75	-0.04
2,4,5-C3A	1.30	1.44	-0.14
2,3,4,5-C4A	1.96	1.83	0.13
2,3,5,6-C4A	1.76	1.66	0.10

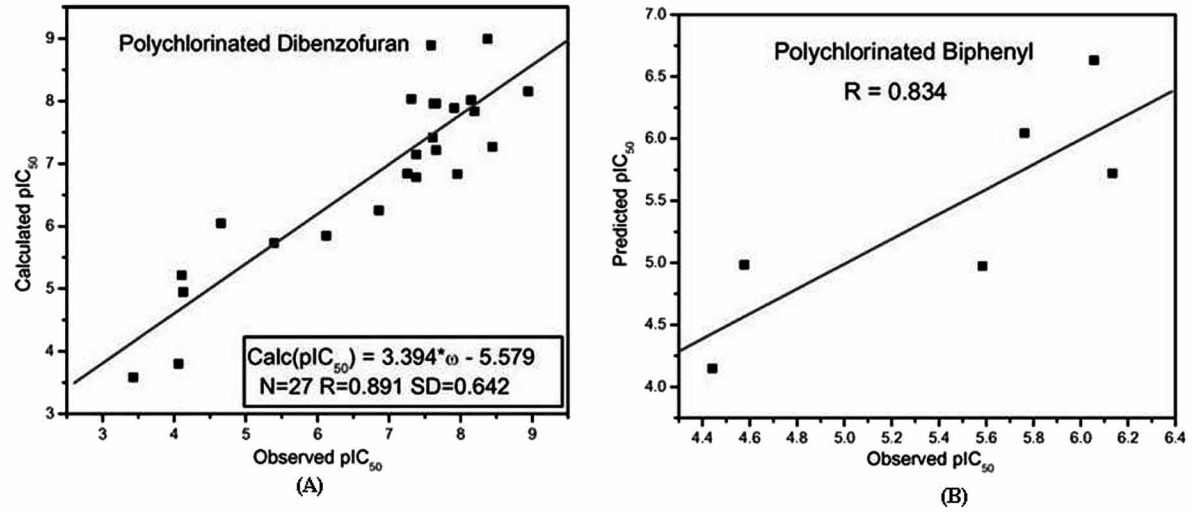
[a] The values from reference 179. [b]Difference between the experimental and calculated toxicity (log(IGC_{50}^{-1})) values. The number before CA indicates the position of chloro group and CxA indicates the presence of x number of chloro groups.

(Reprinted with permission from Padmanabhan et al., 2006a. Copyright © 2006, American Chemical Society).

The corresponding plots of exp. log(IGC_{50}^{-1}) vs. calc. log(IGC_{50}^{-1}) are displayed in Figure 6.

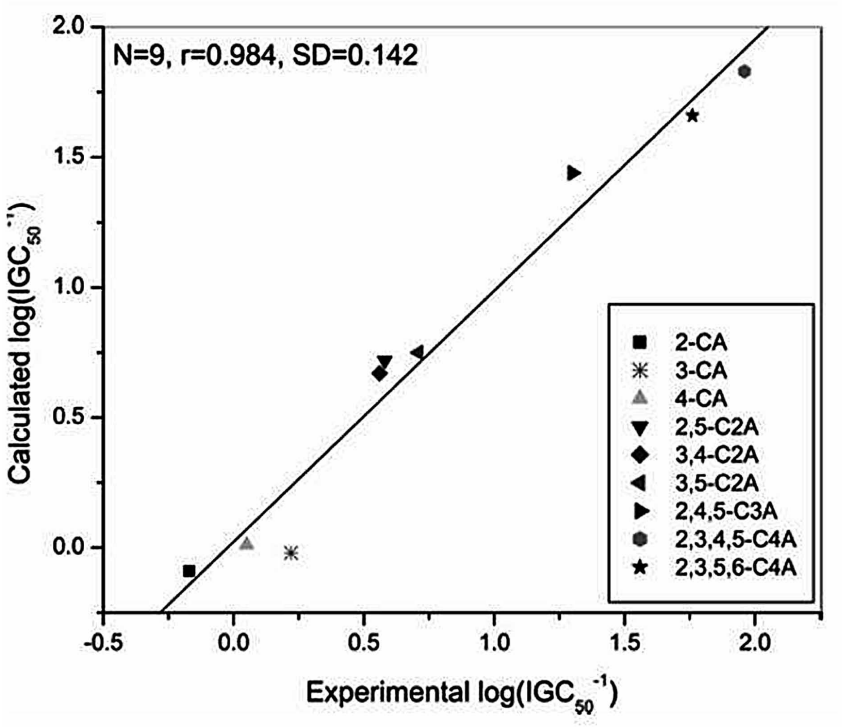
Three-parameter regression models based on global electrophilicity (ω), local philicity descriptor at the O site ($\omega^-_{O\ max}$) in case of phenol derivatives (aromatic donors) and at the C site ($\omega^+_{C\ max}$) in cases of nitrobenzene and benzonitrile derivatives (aromatic acceptors) and total Hartree-Fock energy (E_{HF}) could predict toxicity towards the *T. pyriformis* larger than 80% of cross-validation variance of data of these systems.

Figure 5. A plot between experimental and calculated toxicity (log(IGC_{50}^{-1}) values of the selected chloroanilines against T. pyriformis
(Reprinted with permission from Padmanabhan et al., 2006a. Copyright © 2006, American Chemical Society).



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 6. Experimental versus calculated log (IGC_{50}^{-1}) values for aromatic donors (a) and for aromatic acceptors (b)
(Reprinted with permission from Roy et al., 2006b. Copyright © 2006 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim).



4.6 QSPR Model for the Prediction of Lipophilic Behavior of Polychlorinated Biphenyl Derivatives

Padmanabhan et al. (2006b) successfully developed effective QSPR model for the prediction of the lipophilic behaviour ($logK_{ow}$) of a set having large number (133) of polychlorinated biphenyl (PCB) derivatives by using semi-empirical AM1 method. The effectiveness of a QSPR model depends on the accuracy in the selection of appropriate descriptors. CDFT based global descriptors like ω was employed in addition to the other descriptors viz., energy of the lowest unoccupied molecular orbital (E_{LUMO}) and number of chlorine substituents (N_{Cl}). Assuming experimental $logK_{ow}$ values (Patil, 1991) of the training set of 100 PCB derivatives as dependent variable and these three descriptors along with their several combinations as independent variables, the regression models were developed (see Table 7). It was found that each descriptor individually could describe lipophilic behavior quite well ($r^2 = 0.81$-0.88); however the three-parameter model based on the combination of the ω, E_{LUMO} and N_{Cl} provided the best prediction.

This model was able to predict largest variation in data ($r^2 = 0.914$) with leave-one-out cross-validated squared correlation coefficient ($r^2_{cv} = 0.909$) having lowest SD value of 0.225. Figure 7 depicts the plots between exp. $logK_{ow}$ vs. calc. $logK_{ow}$ for the training set of 100 PCBs and the data set of 133 PCBs. In both cases, the three-parameter regression models gave excellent correlation coefficient, r value of 0.956 for these cases. Therefore, it clearly showed the efficacy of these descriptors in explaining lipophilic behavior of PCBs.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 7. Regression models for logK_{ow} using various descriptors for the training set of 100 PCB congeners

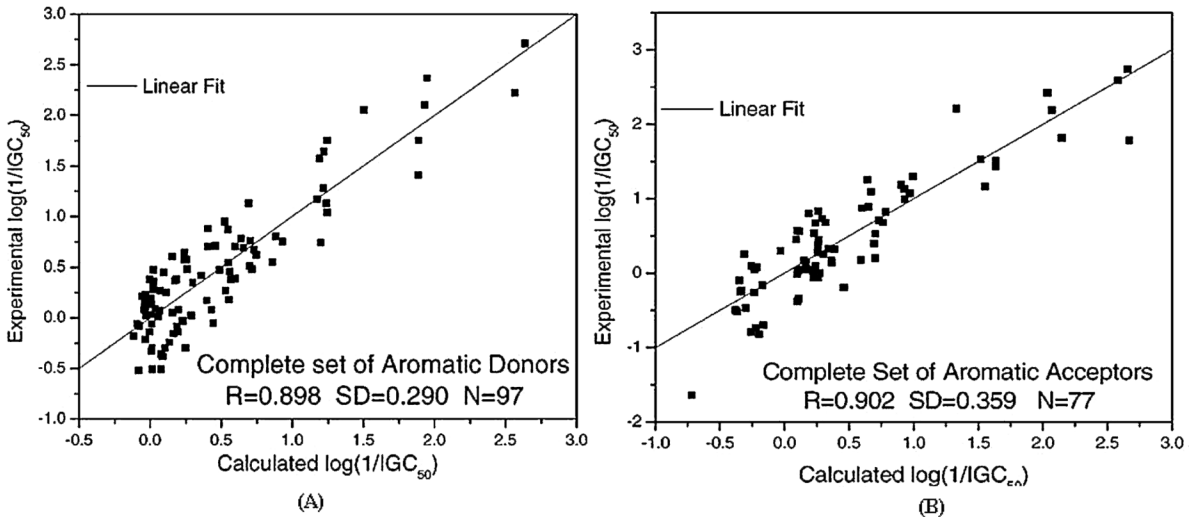
S.No.	Regression Equation	r^2	r^2_{cv}	SD
1	$LogK_{ow} = 4.617–75.219*E_{LUMO}$	0.811	0.806	0.330
2	$LogK_{ow} = 4.234 + 0.406*N_{Cl}$	0.842	0.838	0.302
3	$LogK_{ow} = -2.446 + 81.693*\omega$	0.877	0.874	0.266
4	$LogK_{ow} = 4.262 + 0.243*N_{Cl} - 36.119*E_{LUMO}$	0.892	0.888	0.251
5	$LogK_{ow} = -10.523 + 91.620*E_{LUMO} + 176.280*\omega$	0.905	0.901	0.235
6	$LogK_{ow} = -0.038 + 0.173*N_{Cl} + 51.015*\omega$	0.906	0.902	0.234
7	$LogK_{ow} = -5.967 + 0.115*N_{Cl} + 58.070*E_{LUMO} + 121.280*\omega$	0.914	0.909	0.225

(Reprinted with permission from Padmanabhan et al., 2006b. Copyright © 2006, Elsevier).

4.7 QSPR Models to Predict the Enthalpy of Vaporization of 209 Polychlorinated Biphenyl Congeners

The ω, E_{LUMO} and N_{Cl} were further employed to model QSPR in order to predict the enthalpy of vaporization ($\Delta_{vap}H_m$) of a set of 209 PCBs by using semi-empirical AM1 level of computations (Padmanabhan et al., 2007). A small data set containing only 17 PCBs, whose experimental $\Delta_{vap}H_m$ values were known, (Puri et al., 2001) was considered in building regression models. In general, data set is used to divide into training set and test set but here due to small number of data set, it was only considered as the training set. The different regression models based on the exp. $\Delta_{vap}H_m$ values as dependent parameters and individual ω, E_{LUMO} and N_{Cl} desciptors or their different combinations as the independent parameters along with the obtained r^2, r^2_{cv} and SD values are provided in Table 8.

Figure 7. Experimental versus calculated values of logK_{ow} with electrophilicity, E_{LUMO} and N_{Cl} as descriptors for the model constructed using (a) the training set of 100 PCBs; (b) the data set of 133 PCBs
(Reprinted with permission from Padmanabhan et al., 2006b. Copyright © 2006, Elsevier).



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 8. Regression models for vaporization enthalpy ($\Delta_{vap}H_m$ (298.15 K) in kJ/mol) using various descriptors for the dataset of 17 PCBs[a]

S.No.	Regression Equation	r^2	r^2_{cv}	SD
1	$\Delta_{vap}H_m = 73.8 - 936.5*E_{LUMO}$	0.819	0.764	5.9
2	$\Delta_{vap}H_m = 65.2 + 6.1*N_{Cl}$	0.908	0.879	4.2
3	$\Delta_{vap}H_m = -27.5 + 1159.4*\omega$	0.921	0.904	3.9
4	$\Delta_{vap}H_m = 67.0 + 4.4*N_{Cl} - 309.0*E_{LUMO}$	0.928	0.884	3.9
5	$\Delta_{vap}H_m = 12 + 2.9*N_{Cl} + 654.9*\omega$	0.951	0.918	3.2
6	$\Delta_{vap}H_m = -151.0 + 2583.6*\omega + 1240.4*E_{LUMO}$	0.968	0.943	2.6
7	$\Delta_{vap}H_m = -103.6 + 1.7*N_{Cl} + 994.7*E_{LUMO} + 2009.0*\omega$	0.976	0.948	2.3

[a]Experimental data as obtained from ref 184.

(Reprinted with permission from Padmanabhan et al., 2007. Copyright © 2007 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim).

The regression equation developed by each individual descriptor could predict $\Delta_{vap}H_m$ with high r^2 value (0.82-0.92); however the best correlation was provided by the three-parameter regression equation developed by ω, E_{LUMO} and N_{Cl} desciptors, which described the maximum variation in the data (r^2 = 0.976) having leave-one-out (LOO) cross-validated squared correlation coefficient, r^2_{cv} value of 0.948. This model could explain $\Delta_{vap}H_m$ value with higher efficiency (in terms of higher r^2_{cv}) than that model prescribed by Puri et al. (2001), where they obtained r^2_{cv} value of 0.852. Another advantage of this model was that it could provide excellent result even based on least expensive AM1 level of computations.

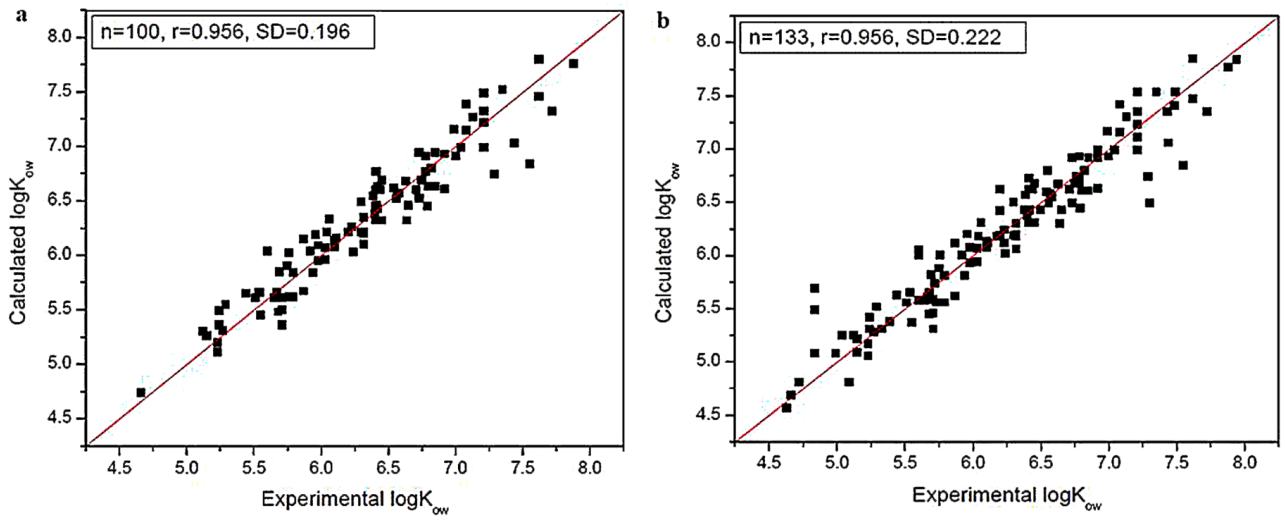
Now, another set of 27 PCDFs with their experimentally reported $\Delta_{vap}H_m$ values (Nakajoh et al., 2006) were considered to further test this model. Here 22 PCBs were taken as training set and remaining five PCBs as test set. It was found that the best regression equation in the previous set could describe maximum variation in the data (r^2 = 0.915) having leave-one-out cross-validated squared correlation coefficient, r^2_{cv} value of 0.856. It was also capable of predicting the $\Delta_{vap}H_m$ values of the test set within 4.2 kJ/mol error bars. Now including these five test set candidates in the data set, for all 27 PCBs, this model was able to represent the $\Delta_{vap}H_m$ values with r^2 = 0.858 and r^2_{cv} = 0.758. The plots between the exp. $\Delta_{vap}H_m$ and calc. $\Delta_{vap}H_m$ values for both the first set and the second set are given in Figure 12. In case of the first set, correlation coefficient (r) was 0.988 and in the second set it was 0.926. Based on this model, the $\Delta_{vap}H_m$ values for the remaining 183 PCB derivatives were predicted (see Figure 8).

4.8 QSAR Model in Predicting Biological Activity of Testosterone and Estrogen Derivatives

Parthasarathi et al. (2004b) further employed electrophilicity to assess its efficacy as a probable biological activity descriptor in modeling QSAR for testosterone and estrogen derivatives. The geometry optimizations of all the systems were carried out at the semi-empirical AM1 level whereas the single point energy calculations at the B3LYP/6-31G(d) level were performed with those geometries obtained at the AM1 level. The biological activity of various testosterone derivatives was described by means of relative binding affinity (RBA) (Saartok et al., 1984), androgenic potency (Counsell et al., 1962), relative androgenic activity (Liao et al., 1973), therapeutic index (Sala & Baldratti, 1957), testosterone-binding globulins (TeBG) binding affinity (Cramer et al., 1988), relative competition indices (Liao et al., 1973),

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 8. Exp. vs. calc. $\Delta_{vap}H_m$ values for the dataset of (a) 17 PCBs, (b) 27 PCBs, in the general set for (c) 193 PCBs and (d) 183 PCBs
(Reprinted with permission from Padmanabhan et al., 2007. Copyright © 2007 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim).



binding affinity for rat ventral prostate receptor protein (Wolff, 1955) and myotrophic to androgenic potency in temporal activity (Kochakian, 1976) (MAPT) whereas the same for estrogen derivatives was measured in terms of RBA values (Gao et al., 1999). In Figure 13, some representative plots between calculated ω and different biological activities of testosterone derivatives are presentated. In every case, a linear correlation between them was found; however, the correlations of RBA, relative competition indices and myotrophic to androgenic potency in temporal activity with ω were found to be better than the others. The corresponding regression models used to generate the plots were inserted within the Figures (see Figure 9). In all cases R^2 values were found to range from 0.61 to 0.90. It was also noted that B3LYP result was better than that of AM1 in predicting biological activity.

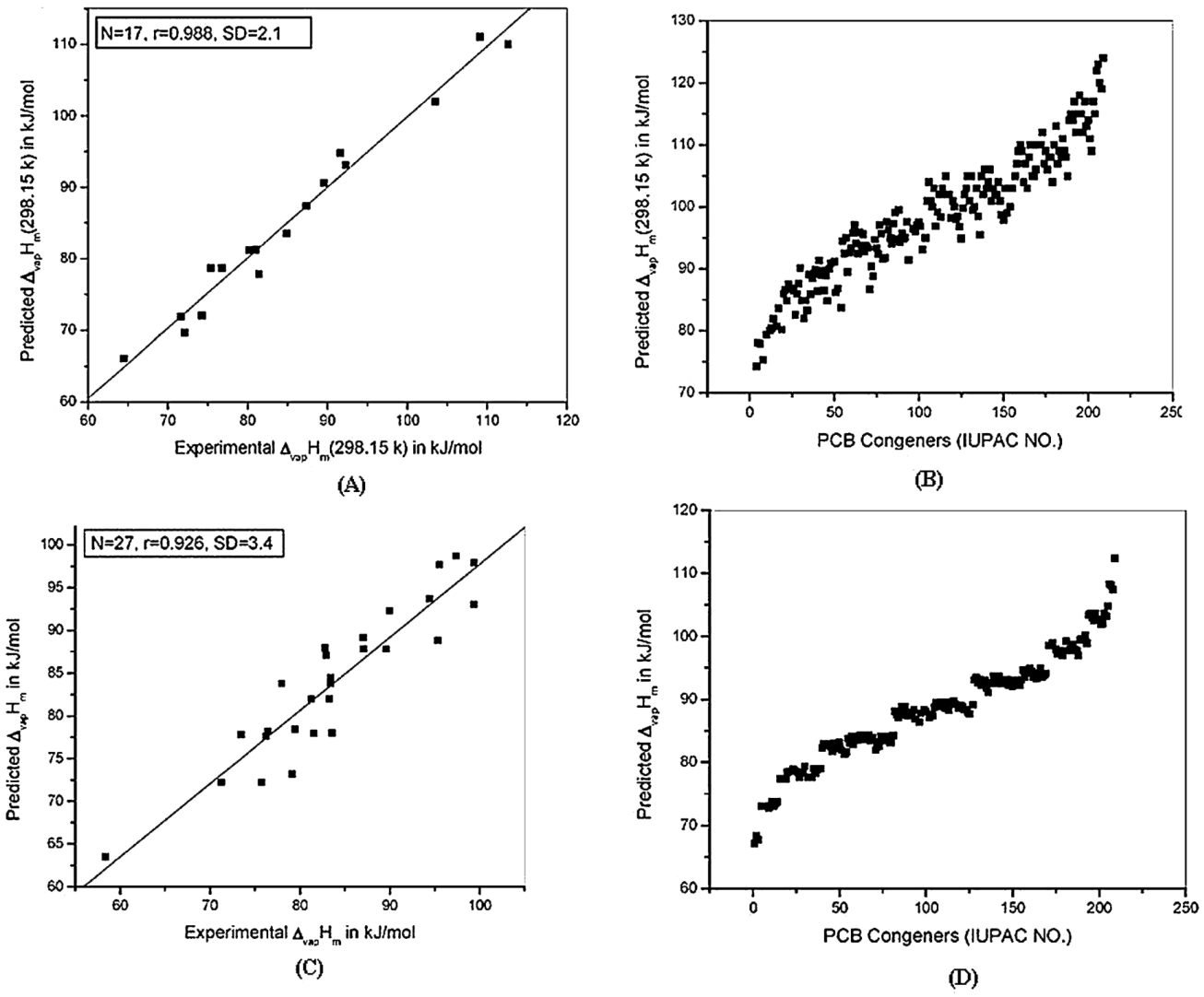
Similarly, in case of estrogen derivatives ω could predict their biological activity towards RBA having R^2 value in the range of 0.73 and 0.87. Therefore, it clearly showed that ω could predict biological activity of these systems quite efficiently.

Roy et al. (2007) also showed that the number of atoms in a molecule (N_A) along with ω values could explain the biological activities of testosterone and estrogen derivatives reasonably well ($R \approx 0.8$).

Pasha et al. (2005) and Singh et al. (2004) developed QSAR models based on frontier orbitals information (E_{HOMO}, E_{LUMO}) and different CDFT based descriptors like η, S, χ, μ as well as different energy descriptors like total energy, heat of formation, electronic energy, and the highest negatively charge on an atom to predict different biological activities of testosterone and estrogen derivatives. They found that the regression models made by η and different energy descriptors possess better predicitivity than the other combinations. They argued that η represents the situation in these cases in a better way due to the associated maximum hardness principle. Srivastava et al. (2005) further showed that some other

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 9. Relationship between various biological activities of testosterone derivatives with electrophilicity index
(Reprinted with permission from Parthasarathi et al., 2004b. Copyright © 2004, Elsevier).



descriptors like IP and EA of an atom in a molecule, atomic softness could predict the biological activity of testosterone derivatives quite well, The differences in softness values of estrogen derivatives and lysine, histidine, tyrosine and cysteine receptors were also found to be an effective parameter in describing the biological activities of estrogen derivatives (Pasha et al., 2005). Rokhina & Suri (2012) further built QSPR to represent the estrogenicity of the natural estrogen hormones. They found that the information of frontier molecular orbital energies is very important to interpret this activity.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

5. GROUP PHILICITY TOWARDS MODELING QSAR

In addition to the conventional global electrophilicity Padmanabhan et al. (2006c) used group philicity (ω_g^+, ω_g^-) computed at the B3LYP/6-31G(d) level to model QSAR for ecotoxicity of chlorophenols (CPs). The data set was made by 18 CPs, which was divided into training set having 16 CPs and test set containing two CPs. Then various regression models were developed by these descriptors in order to describe the experimental toxicity ($\log(1/EC_{50})$) (Devillers & Chambon, 1986) of the training set against *Daphnia magna* taking $\log(1/EC_{50})$ as a dependent variable whereas ω, ω_g^+ and ω_g^- as independent variables (see Table 9). It was found that two-parameter regression equation made by ω_g^+ and ω_g^- was capable of describing maximum variation data (88.7%) having large r^2_{cv} value of 0.826. The SD was also found to be the least for this model. This model was also found to predict the experimental $\log(1/EC_{50})$ value for the test set within $\pm$ 0.05 error bar. Now, regression analyses were carried out for all 18 CPs (data set) (see Table 9). In this set also, the combination of ω_g^+ and ω_g^- could explain maximum variation data (88.9%) along with r^2_{cv} value of 0.838. The second best model was three-parameter regression composed by ω, ω_g^+ and ω_g^-, which could also predict 88.9% variation data, however, the r^2_{cv} value for this model was 0.812. The plots of calc. $\log(1/EC_{50})$ vs exp. $\text{Log}(1/EC_{50})$ are given in Figure 10(A), which shows a correlation coefficient of 0.942 for the best model. Further, these descriptors were used to explain the experimental toxicity of CPs against *B. rerio*, and *Bacillus*. It was found that in case against *B. rerio*, three-parameter regression model (ω, ω_g^+ and ω_g^-), being a very good model, could explain

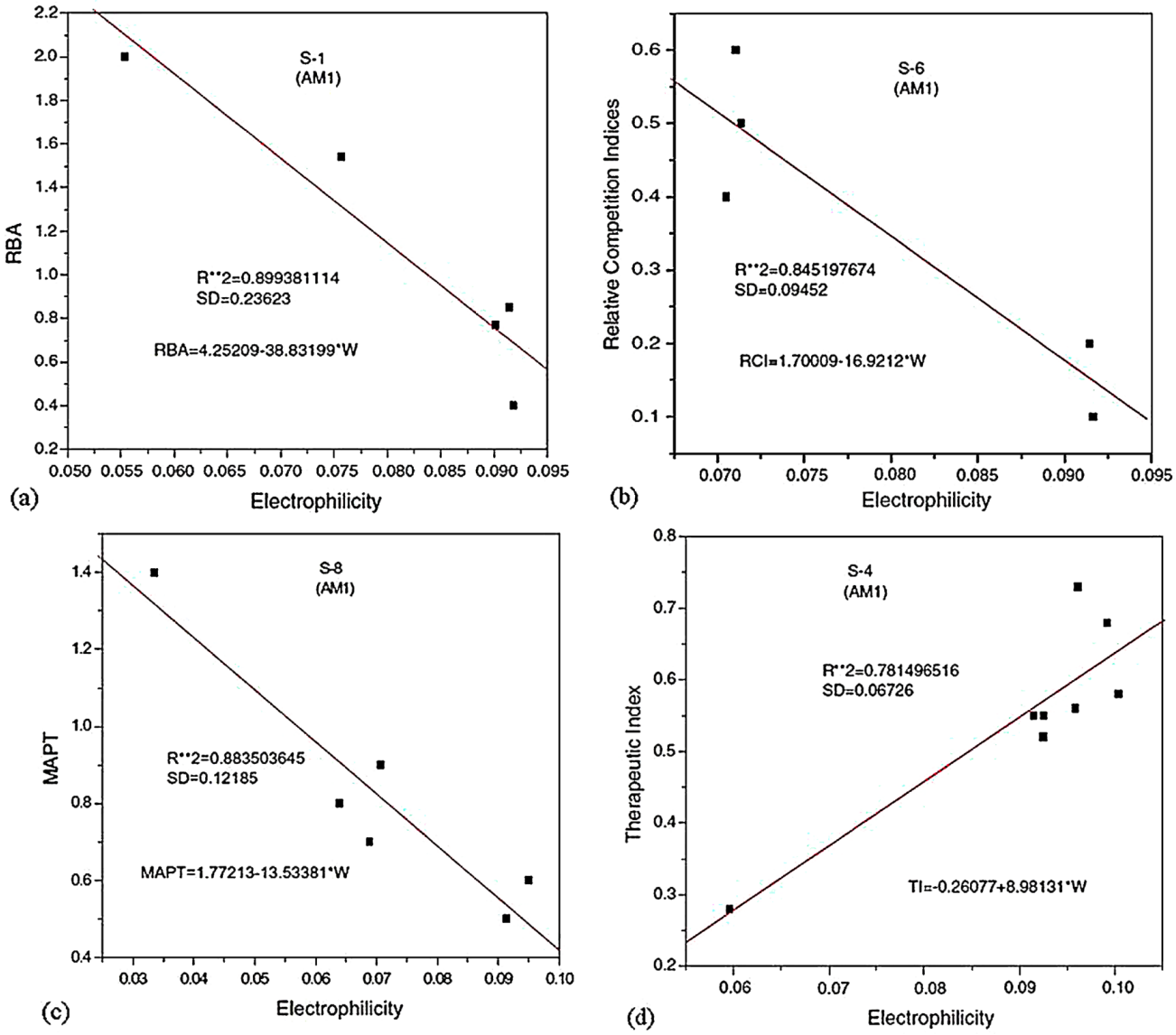
Table 9. Regression models for the toxicity [$\log(1/EC_{50})$] against D. magna using various descriptors for a training Set of 16 CPs and data set of 18 CPs

S.No.	Regression Equation	r^2	r^2_{cv}	SD
	For training set			
1	$\log(1/EC_{50}) = -2.2 + 198.0*\omega_g^-$	0.515	0.374	0.348
2	$\log(1/EC_{50}) = -0.9 + 28.7*\omega$	0.731	0.671	0.259
3	$\log(1/EC_{50}) = -1.4 + 24.8*\omega + 42.0*\omega_g^-$	0.741	0.630	0.264
4	$\log(1/EC_{50}) = -3.8 + 545.6*\omega_g^+$	0.874	0.835	0.177
5	$\log(1/EC_{50}) = -3.5 + 3.5*\omega + 491.3*\omega_g^+$	0.877	0.802	0.178
6	$\log(1/EC_{50}) = -4.0 - 0.6*\omega + 492.3*\omega_g^+ + 43.26*\omega_g^-$	0.887	0.792	0.182
7	$\log(1/EC_{50}) = -4.0 + 485.3*\omega_g^+ + 41.9*\omega_g^-$	0.887	0.826	0.175
	For data set			
1	$\log(1/EC_{50}) = -2.2 + 196.2*\omega_g^-$	0.529	0.409	0.326
2	$\log(1/EC_{50}) = -0.8 + 28.4*\omega$	0.704	0.644	0.260
3	$\log(1/EC_{50}) = -1.6 + 22.4*\omega + 65.5*\omega_g^-$	0.731	0.634	0.254
4	$\log(1/EC_{50}) = -3.8 + 545.5*\omega_g^+$	0.876	0.842	0.167
5	$\log(1/EC_{50}) = -3.6 + 3.3*\omega + 496.4*\omega_g^+$	0.879	0.817	0.171
6	$\log(1/EC_{50}) = -3.9 - 0.6*\omega + 477.3*\omega_g^+ + 40.2*\omega_g^-$	0.889	0.812	0.170
7	$\log(1/EC_{50}) = -4.0 + 484.6*\omega_g^+ + 41.4*\omega_g^-$	0.889	0.838	0.164

(Reprinted with permission from Padmanabhan et al., 2006c. Copyright © 2006, American Chemical Society).

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 10. Plots between experimental and calculated toxicity of CPs against (A) D. magna, (B) B. rerio, and (C) Bacillus
(Reprinted with permission from Padmanabhan et al., 2006c. Copyright © 2006, American Chemical Society).



92.8% variation data ($r^2_{cv} = 0.878$) having correlation coefficient value (r) of 0.964 (see Figure 10(B)) whereas in case against *Bacillus* this model was able to describe maximum variation in data (91.9%) along with r^2_{cv} value of 0.863 and r value of 0.959 (see Figure 10(C).

Therefore, this result implied that group philicity along with electrophilicity are very much effective in representing the toxicity of CPs against *D. magna*, *B. rerio*, and *Bacillus*. It in turn showed the importance of these reactivity indices as effective descriptors in ecotoxicological studies.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

6. EFFECT OF CHOSEN FUNCTIONALS IN BUILDING GLOBAL REACTIVITY DESCRIPTORS BASED QSAR/QSPR/QSTR MODELS

Vijayaraj et al. (2009) assessed the effect of studied functionals on the ionization potential (*IP*) and the electron affinity (*EA*) of the systems, which were generally used to compute the CDFT based reactivity descriptors. Therefore, the dependence of *IP* and *EA* on the chosen functionals would raise questions about the level of theory chosen to construct QSAR, QSPR and QSTR based on the reactivity descriptors. In other words, the predictive power of a regression model might depend on the chosen level of theory. In order to explore the efficacy of various density functionals along with different basis sets in calculating *IP* and *EA*, hence, to benchmark, systematic computations were performed considering a series of chlorinated benzenes (CBs). Some common functionals of pure, GGA, hybrid, and meta and hybrid meta GGA types were employed. It was found that a good agreement between computed and experimental *IP* and E*A* values could be attained with the increase in the rigor of the level of computation. The prediction of *IP* and *EA* of CBs, computed using both Koopmans' approximation and ΔSCF method, was most accurately made by the use of M05-2X functional. Further, in between Koopmans' approximation and ΔSCF method, *IP* and *EA* computed by ΔSCF method matched more closely with the corresponding experimental values than that by the Koopmans' approximation.

Now, using those descriptors computed at different levels of theory, different regression models were designed to predict the toxicity of nine CBs against *Poecilia reticulata*. Different descriptors along with different approximations were included. It was found that although the computation of global reactivity descriptors at different levels and/or different basis sets yielded considerably different values, the mean deviation of a certain property among all the studied systems with respect to the chosen method/basis set was comparable. It meant that with the change in studied method and/or basis set, the values of computed global reactivity descriptors would be changed, however, the magnitude of the difference in between the systems were more or less same in all the methods. This finding was important in this sense that the efficacy of a regression model and hence the correlation coefficient would not alter significantly with the change in level of theory. For example, in spite of good agreement with the experimental *IP* and *EA* using ΔSCF method, it did not improve the correlation coefficient with respect to that obtained using Koopmans' theorem. In general, to construct a QSAR/QSPR/QSTR model, one has to consider a data set containing a large number of systems. Therefore, it can be concluded that if one considers computationally a very cheap level of theory to model QSAR/QSPR/QSTR, it can be possible to get reliable results similar to the one constructed from the results of a high level of theory, which will in deed save a lot of computational time.

7. CONCLUSION

In order to better understand the variation of property, activity and toxicity with the structural change of a molecular motif, the Quantitative-Structure-(Property/Activity/Toxicity) Relationship models are widely applied. Various types of methods and models were employed to construct effective regression equations. Both, global and local reactivity descriptors based on Conceptual Density Functional Theory are found to be very effective indices in order to predict biological activity or property or toxicity of a set of systems. We have specially drawn attention on the efficacy of the regression models built

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

based on the different combinations of the parameters viz., electrophilicity index, net electrophilicity index, group philicity along with their local variants, atomic charge, number of carbon or Cl or non-hydrogenic atoms, atomic number, chemical potential, LUMO energy and Hartree-Fock energy in predicting the toxicity of various aliphatic compounds, polyaromatic hydrocarbons, chloroanilines, phenol, nitrobenzene, benzonitrile-derivatives, halogen, sulfur and chlorinated aromatic compounds, testosterone, estrogen derivatives, and different alkali, alkaline-earth, transition-metal, arsenic ions as well. Although individual descriptors were found to vary with the change in level of theory and /or basis set, the efficacy of the overall QSAR/QSPR/QSTR models depend hardly on the level of theory. Therefore, very low level of theory can also yield reliable results.

Therefore, the reactivity descriptors within the purview of conceptual DFT will be a good choice in predicting the desired properties of a set of systems.

ACKNOWLEDGMENT

We are grateful to Professor Kunal Roy for inviting us to contribute this chapter. PKC would like to thank DST, New Delhi for the J. C. Bose National Fellowship. SP thanks CSIR, New Delhi for his fellowship.

REFERENCES

Aboye, T. L., Sobhia, M. E., & Bharatam, P. V. (2004). 3D-QSAR studies of pyruvate dehydrogenase kinase inhibitors based on a divide and conquer strategy. *Bioorganic & Medicinal Chemistry*, *12*(10), 2709–2715. doi:10.1016/j.bmc.2004.03.019 PMID:15110852

Abraham, M. H. (1994). New solute descriptors for linear free energy relationships and quantitative structure-activity relationships. In Quantitative treatments of solute/solvent interactions. Elsevier.

Abraham, M. H., Chadha, H. S., Dixon, J. P., Rafols, C., & Treiner, C. (1995). Hydrogen bonding. Part 40. Factors that influence the distribution of solutes between water and sodium dodecylsulfate micelles. *Journal of the Chemical Society, Perkin Transactions*, *2*(5), 887–894. doi:10.1039/p29950000887

Adhikari, N., Halder, A. K., Mondal, C., & Jha, T. (2013). Exploring structural requirements of aurone derivatives as antimalarials by validated DFT-based QSAR, HQSAR, and COMFA–COMSIA approach. *Medicinal Chemistry Research*, *22*(12), 6029–6045. doi:10.1007/s00044-013-0590-8

Adhikari, N., Halder, A. K., Mondal, C., & Jha, T. (2014). Structural findings of quinolone carboxylic acids in cytotoxic, antiviral, and anti-HIV-1 integrase activity through validated comparative molecular modeling studies. *Medicinal Chemistry Research*, *23*(6), 3096–3127. doi:10.1007/s00044-013-0897-5

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

Amin, E. A., & Welsh, W. J. (2006). Highly predictive CoMFA and CoMSIA models for two series of stromelysin-1 (MMP-3) inhibitors elucidate S1‘ and S1−S2‘ binding modes. *Journal of Chemical Information and Modeling*, *46*(4), 1775–1783. doi:10.1021/ci060089d PMID:16859309

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Arning, J., Stolte, S., Böschen, A., Stock, F., Pitner, W.-R., & Welz-Biermann, U. et al. (2008). Qualitative and quantitative structure activity relationships for the inhibitory effects of cationic head groups, functionalised side chains and anions of ionic liquids on acetylcholinesterase. *Green Chemistry*, *10*(1), 47–58. doi:10.1039/b712109a

Avery, M. A., Alvim-Gaston, M., Rodrigues, C. R., Barreiro, E. J., Cohen, F. E., Sabnis, Y. A., & Woolfrey, J. R. (2002). Structure−activity relationships of the antimalarial agent artemisinin. 6. The development of predictive in vitro potency models using CoMFA and HQSAR methodologies. *Journal of Medicinal Chemistry*, *45*(2), 292–303. doi:10.1021/jm0100234 PMID:11784134

Ayers, P. W., & Levy, M. (2000). Perspective on "Density functional approach to the frontier-electron theory of chemical reactivity". *Theoretical Chemistry Accounts*, *103*(3-4), 353–360. doi:10.1007/s002149900093

Ayers, P. W., & Parr, R. G. (2000). Variational principles for describing chemical reactions: The Fukui function and chemical hardness revisited. *Journal of the American Chemical Society*, *122*(9), 2010–2018. doi:10.1021/ja9924039

Bajot, F. (2010). The use of QSAR and computational methods in drug design. In *Recent advances in QSAR studies* (pp. 261–282). Springer. doi:10.1007/978-1-4020-9783-6_9

Balaban, A. T. (1997). From chemical topology to 3D geometry. *Journal of Chemical Information and Computer Sciences*, *37*(4), 645–650. doi:10.1021/ci960168x

Balaban, A. T. (1998). Topological and stereochemical molecular descriptors for databases useful in QSAR, similarity/dissimilarity and drug design. *SAR and QSAR in Environmental Research*, *8*(1-2), 1–21. doi:10.1080/10629369808033259

Bao, Y., Huang, Q., Li, Y., Li, N., He, T., & Feng, C. (2012). Prediction of nitrobenzene toxicity to the algae (Scenedesmus obliguus) by quantitative structure–toxicity relationship (QSTR) models with quantum chemical descriptors. *Environmental Toxicology and Pharmacology*, *33*(1), 39–45. doi:10.1016/j.etap.2011.09.003 PMID:22088683

Bearden, A. P., & Schultz, T. W. (1998). Comparison of *Tetrahymena* and *Pimephales* toxicity based on mechanism of action. *SAR and QSAR in Environmental Research*, *9*(3-4), 127–153. doi:10.1080/10629369808039153 PMID:9933957

Berkowitz, M., & Parr, R. G. (1988). Molecular hardness and softness, local hardness and softness, hardness and softness kernels, and relations among these quantities. *The Journal of Chemical Physics*, *88*(4), 2554–2557. doi:10.1063/1.454034

Bhongade, B. A., & Gadad, A. K. (2004). 3D-QSAR CoMFA/CoMSIA studies on urokinase plasminogen activator (uPA) inhibitors: A strategic design in novel anticancer agents. *Bioorganic & Medicinal Chemistry*, *12*(10), 2797–2805. doi:10.1016/j.bmc.2004.02.019 PMID:15110861

Blaney, J., & Hansch, C. (1990). Application of molecular graphics to the analysis of macromolecular structure. *Comprehensive Medicinal Chemistry*, *4*, 459–496.

Blankley, C. J. (1996). *Structure property correlations in drug research*. Academic Press.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Bohari, M. H., Srivastava, H. K., & Sastry, G. N. (2011). Analogue-based approaches in anti-cancer compound modelling: The relevance of QSAR models. *Organic and Medicinal Chemistry Letters*, *1*(1), 1–12. doi:10.1186/2191-2858-1-3 PMID:22373294

Bonaccorsi, R., Scrocco, E., & Tomasi, J. (1970). Molecular SCF calculations for the ground state of some three-membered ring molecules: $(CH_2)_3$, $(CH_2)_2NH$, $(CH_2)NH_2^+$, $(CH_2)_2O$, $(CH_2)_2S$, $(CH)_2CH_2$, and N_2CH_2. *The Journal of Chemical Physics*, *52*(10), 5270–5284. doi:10.1063/1.1672775

Buttingsrud, B., Alsberg, B. K., & Åstrand, P.-O. (2007a). An investigation of descriptors based on the critical points in the electron density by building quantitative structure–property relationships for proton chemical shifts. *Journal of Molecular Structure THEOCHEM*, *810*(1-3), 15–24. doi:10.1016/j.theochem.2007.01.031

Buttingsrud, B., Alsberg, B. K., & Åstrand, P.-O. (2007b). Quantitative prediction of the absorption maxima of azobenzene dyes from bond lengths and critical points in the electron density. *Physical Chemistry Chemical Physics*, *9*, 2226–2233. doi:10.1039/b617470a PMID:17487319

Buttingsrud, B., Alsberg, B. K., & Åstrand, P.-O. (2007c). Validation of critical points in the electron density as descriptors by building quantitative structure-property relationships for the atomic polar tensor. *Journal of Computational Chemistry*, *28*(13), 2130–2139. doi:10.1002/jcc.20666 PMID:17464968

Camenisch, G., Folkers, G., & van de Waterbeemd, H. (1998). Shapes of membrane permeabilitylipophilicity curves: Extension of theoretical models with an aqueous pore pathway. *European Journal of Pharmaceutical Sciences*, *6*(4), 321–329. doi:10.1016/S0928-0987(98)00033-5

Chakraborty, A., Duley, S., Giri, S., & Chattaraj, P. K. (2012). An understanding of the origin of chemical reactivity from a conceptual DFT approach. In J. Wiley (Ed.), A matter of density, exploring the electron density concept in the chemical, biological, and materials sciences. Academic Press. doi:10.1002/9781118431740.ch7

Chakraborty, A., Pan, S., & Chattaraj, P. K. (2013). Biological activity and toxicity: A conceptual DFT approach. *Structure and Bonding*, *150*, 143–179. doi:10.1007/978-3-642-32750-6_5

Chamorro, E., Chattaraj, P. K., & Fuentealba, P. (2003). Variation of the electrophilicity index along the reaction path. *The Journal of Physical Chemistry A*, *107*(36), 7068–7072. doi:10.1021/jp035435y

Chattaraj, P. K. (1992). Electronegativity and hardness: A density functional treatment. *Journal of the Indian Chemical Society*, *69*, 173–183.

Chattaraj, P. K. (1996). The maximum hardness principle: An overview. *Proceedings of the Indian National Science Academy - Part A*, *62*, 513–531.

Chattaraj, P. K. (Ed.). (2009). *Chemical reactivity theory: A density functional view*. Taylor and Francis/CRC Press. doi:10.1201/9781420065442

Chattaraj, P. K., Arun Murthy, T. V. S., Giri, S., & Roy, D. R. (2007). A connection between softness and magnetizability. *Journal of Molecular Structure THEOCHEM*, *813*(1-3), 63–65. doi:10.1016/j.theochem.2007.02.035

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chattaraj, P. K., Chakraborty, A., & Giri, S. (2009). Net electrophilicity. *The Journal of Physical Chemistry A*, *113*(37), 10068–10074. doi:10.1021/jp904674x PMID:19702288

Chattaraj, P. K., Fuentealba, P., Gomez, B., & Contreras, R. (2000). Woodward-Hoffmann rule in the light of the principles of maximum hardness and minimum polarizability: DFT and ab initio SCF studies. *Journal of the American Chemical Society*, *122*(2), 348–351. doi:10.1021/ja992337a

Chattaraj, P. K., Fuentealba, P., Jaque, P., & Toro-Labbé, A. (1999). Validity of the minimum polarizability principle in molecular vibrations and internal rotations: An ab initio SCF study. *The Journal of Physical Chemistry A*, *103*(46), 9307–9312. doi:10.1021/jp9918656

Chattaraj, P. K., & Giri, S. (2009). Electrophilicity index within a conceptual DFT framework. *Annual Reports on the Progress of Chemistry Section C*, *105*, 13–39. doi:10.1039/b802832j

Chattaraj, P. K., Giri, S., & Duley, S. (2011). Update 2 of: Electrophilicity index. *Chemical Reviews*, *111*(2), PR43–PR75. doi:10.1021/cr100149p PMID:21306180

Chattaraj, P. K., Liu, G. H., & Parr, R. G. (1995). The maximum hardness principle in the Gyftopoulos-Hatsopoulos three-level model for an atomic or molecular species and its positive and negative ions. *Chemical Physics Letters*, *237*(1-2), 171–176. doi:10.1016/0009-2614(95)00280-H

Chattaraj, P. K., Maiti, B., & Sarkar, U. (2003). Philicity: A unified treatment of chemical reactivity and selectivity. *The Journal of Physical Chemistry A*, *107*(25), 4973–4975. doi:10.1021/jp034707u

Chattaraj, P. K., & Poddar, A. (1998). A density functional treatment of chemical reactivity and the associated electronic structure principles in the excited electronic states. *The Journal of Physical Chemistry A*, *102*(48), 9944–9948. doi:10.1021/jp982734s

Chattaraj, P. K., & Roy, D. R. (2007). Update 1 of: Electrophilicity index. *Chemical Reviews*, *107*(9), PR46–PR74. doi:10.1021/cr078014b PMID:21306180

Chattaraj, P. K., Roy, D. R., Geerlings, P., & Torrent-Sucarrat, M. (2007). Local hardness: A critical account. *Theoretical Chemistry Accounts*, *118*(5-6), 923–930. doi:10.1007/s00214-007-0373-8

Chattaraj, P. K., Roy, D. R., Giri, S., Mukherjee, S., Subramanian, V., & Parthasarathi, R. et al. (2007). An atom counting and electrophilicity based QSTR approach. *Journal of Chemical Sciences*, *119*(5), 475–488. doi:10.1007/s12039-007-0061-1

Chattaraj, P. K., Sarkar, U., & Roy, D. R. (2006). Electrophilicity index. *Chemical Reviews*, *106*(6), 2065–2091. doi:10.1021/cr040109f PMID:16771443

Chattaraj, P. K., & Sengupta, S. (1996). Popular electronic structure principles in a dynamical context. *Journal of Physical Chemistry*, *100*(40), 16126–16130. doi:10.1021/jp961096f

Chattaraj, P. K., & Sengupta, S. (1997). Dynamics of chemical reactivity indices for a many-electron system in its ground and excited states. *The Journal of Physical Chemistry A*, *101*(42), 7893–7900. doi:10.1021/jp971408u

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chaudry, U. A., & Popelier, P. L. A. (2003). Ester hydrolysis rate constant prediction from quantum topological molecular similarity descriptors. *The Journal of Physical Chemistry A*, *107*(22), 4578–4582. doi:10.1021/jp034272a

Chaudry, U. A., & Popelier, P. L. A. (2004). Estimation of pKa using quantum topological molecular similarity descriptors: application to carboxylic acids, anilines and phenols. *The Journal of Organic Chemistry*, *69*(2), 233–241. doi:10.1021/jo0347415 PMID:14725434

Consonni, V., Todeschini, R., & Pavan, M. (2002). Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. Part 1. Theory of the novel 3D molecular descriptors. *Journal of Chemical Information and Computer Sciences*, *42*(3), 682–692. doi:10.1021/ci015504a PMID:12086530

Counsell, R. E., Klimstra, P. D., & Colton, F. B. (1962). Anabolic agents. Derivatives of 5α-Androst-1-ene. *The Journal of Organic Chemistry*, *27*(1), 248–253. doi:10.1021/jo01048a060

Cramer, R. D. III, & Milne, M. (1979). Abstracts of papers of the American Chemical Society. *Computer Chemistry Section, 44*.

Cramer, R. D. III, 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., Gregory, B. W., & Schultz, T. W. (1998). Quantitative structure-activity analyses of nitrobenzene toxicity to tetrahymena pyriformis. *Chemical Research in Toxicology*, *11*(8), 902–908. doi:10.1021/tx970166m PMID:9705752

Cronin, M. T. D., & Schultz, T. W. (1996). Structure-toxicity relationships for phenols to *Tetrahymena pyriformis. Chemosphere*, *32*(8), 1453–1468. doi:10.1016/0045-6535(96)00054-9 PMID:8653384

Cronin, M. T. D., & Schultz, T. W. (1998). Structure–toxicity relationships for three mechanisms of action of toxicity to *vibrio fischeri. Ecotoxicology and Environmental Safety*, *39*(1), 65–69. doi:10.1006/eesa.1997.1618 PMID:9515077

Cruciani, G., Pastor, M., & Guba, W. (2000). VolSurf: A new tool for the pharmacokinetic optimization of lead compounds. *European Journal of Pharmaceutical Sciences*, *11*, S29–S39. doi:10.1016/S0928-0987(00)00162-7 PMID:11033425

Dalgarno, A. (1962). Atomic polarizabilities and shielding factors. *Advances in Physics*, *11*(44), 281–315. doi:10.1080/00018736200101302

Damme, S. V., & Bultinck, P. (2010). 3D QSAR based on conceptual DFT molecular fields: Antituberculotic activity. *Journal of Molecular Structure THEOCHEM*, *943*(1-3), 83–89. doi:10.1016/j.theochem.2009.10.031

de Melo, E. B., & Ferreira, M. M. C. (2009). Multivariate QSAR study of 4,5-dihydroxypyrimidine carboxamides as HIV-1 integrase inhibitors European. *Journal of Medicinal Chemistry*, *44*(9), 3577–3583. doi:10.1016/j.ejmech.2009.03.001 PMID:19327872

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Debnath, A. K., Shusterman, A. J., Lopez de Compadre, R. L., & Hasch, C. (1994). The importance of the hydrophobic interaction in the mutagenicity of organic compounds. *Mutation Research/Fundamental and Molecular Mechanisms, 305*, 63-72.

Devillers, J., & Chambon, P. (1986). Acute toxicity and QSAR of chlorophenols on *Daphnia magna*. *Bulletin of Environmental Contamination and Toxicology*, *37*(1), 599–605. doi:10.1007/BF01607810 PMID:3768559

Eroglu, E., Palaz, S., Oltulu, O., Turkmen, H., & Ozaydın, C. (2007). Comparative QSTR study using semi-empirical and first principle methods based descriptors for acute toxicity of diverse organic compounds to the fathead minnow. *International Journal of Molecular Sciences*, *8*(12), 1265–1283. doi:10.3390/ijms8121265

Ferguson, A. M., Heritage, T. W., Jonathon, P., Pack, S. E., Phillips, L., Rogan, J., & Snaith, P. J. (1997). EVA: A new theoretically based molecular descriptor for use in QSAR\QSPR analysis. *Journal of Computer-Aided Molecular Design*, *11*(2), 143–152. doi:10.1023/A:1008026308790 PMID:9089432

Filho, E. B. A., Moraes, I. A., Weber, K. C., Rocha, G. B., & Vasconcellos, M. L. A. A. (2012). DFT/PCM, QTAIM, 1H NMR conformational studies and QSAR modeling of thirty-two anti-Leishmania amazonensis Morita–Baylis–Hillman Adducts. *Journal of Molecular Structure THEOCHEM*, *1022*, 72–80. doi:10.1016/j.molstruc.2012.04.051

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

Fuentealba, P., & Reyes, O. (1993). Atomic softness and the electric dipole polarizability. *Journal of Molecular Structure THEOCHEM*, *282*(1-2), 65–70. doi:10.1016/0166-1280(93)85035-W

Fujita, T., & Ban, T. (1971). Structure-activity relation. 3. 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

Fukui, K. (1975). *Theory of orientation and stereoselection*. Berlin: Springer. doi:10.1007/978-3-642-61917-5

Fukui, K. (1982). Role of frontier orbitals in chemical reactions. *Science*, *218*(4574), 747–754. doi:10.1126/science.218.4574.747 PMID:17771019

Fukui, K., Yonezawa, T., Nagata, C., & Shingu, H. (1954). Molecular orbital theory of orientation in aromatic, heteroaromatic, and other conjugated molecules. *The Journal of Chemical Physics*, *22*(8), 1433–1442. doi:10.1063/1.1740412

Fukui, K., Yonezawa, T., & Shingu, H. (1952). A molecular orbital theory of reactivity in aromatic hydrocarbons. *The Journal of Chemical Physics*, *20*(4), 722–725. doi:10.1063/1.1700523

Gao, H., Katzenellenbogen, J. A., Garg, R., & Hansch, C. (1999). Comparative QSAR analysis of estrogen receptor ligands. *Chemical Reviews*, *99*(3), 723–744. doi:10.1021/cr980018g PMID:11749430

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Garg, R., Gupta, S. P., Gao, H., Babu, M. S., Debnath, A. K., & Hansch, C. (1999). Comparative quantitative structure-activity relationship studies on anti-HIV drugs. *Chemical Reviews*, *99*(12), 3525–3602. doi:10.1021/cr9703358 PMID:11849030

Gázquez, J. L., Cedillo, A., & Vela, A. (2007). Electrodonating and electroaccepting powers. *The Journal of Physical Chemistry A*, *111*(10), 1966–1970. doi:10.1021/jp065459f PMID:17305319

Geerlings, P., De. Proft, F., & Langenaeker, W. (2003). Conceptual density functional theory. *Chemical Reviews*, *103*(5), 1793–1874. doi:10.1021/cr990029p PMID:12744694

Ghanty, T. K., & Ghosh, S. K. (1993). Correlation between hardness, polarizability, and size of atoms, molecules, and clusters. *Journal of Physical Chemistry*, *97*(19), 4951–4953. doi:10.1021/j100121a015

Ghanty, T. K., & Ghosh, S. K. (1996). A density functional approach to hardness, polarizability, and valency of molecules in chemical reactions. *Journal of Physical Chemistry*, *100*(30), 12295–12298. doi:10.1021/jp960276m

Ghosh, S. K., & Berkowitz, M. J. (1985). A classical fluid-like approach to the density-functional formalism of many-electron systems. *The Journal of Chemical Physics*, *83*(6), 2976–2983. doi:10.1063/1.449846

Gironés, X., & Carbó-Dorca, R. (2006). Modelling Toxicity using Molecular Quantum Similarity Measures. *QSAR & Combinatorial Science*, *25*(7), 579–589. doi:10.1002/qsar.200530128

Gombar, V. K., Mattioni, B. E., Zwickl, C., & Deahl, J. T. (2006) Computational approaches for assessment of toxicity: A historical perspective and current status. In Computational toxicology: Risk assessment for pharmaceutical and environmental chemicals. John Wiley & Sons, Inc. doi: (Chapter 7). doi:10.1002/9780470145890

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

Griffith, D. J. (1999). *Introduction to electrodynamics*. Prentice Hall, Inc.

Grover, M., Singh, B., Bakshi, M., & Singh, S. (2000a). Quantitative structure–Property relationships in pharmaceutical research – Part 1. *Pharmaceutical Science & Technology Today*, *3*, 28–35.

Grover, M., Singh, B., Bakshi, M., & Singh, S. (2000b). Quantitative structure–property relationships in pharmaceutical research – Part 2. *Pharmaceutical Science & Technology Today*, *3*(2), 50–57. doi:10.1016/S1461-5347(99)00215-1 PMID:10664573

Gyftopoulos, E. P., & Hatsopoulos, G. N. (1968). Quantum-thermodynamic definition of electronegativity. *Proceedings of the National Academy of Sciences of the United States of America*, *60*(3), 786–793. doi:10.1073/pnas.60.3.786 PMID:16591659

Haghdadi, M., & Fatemi, M. H. (2010). Quantitative structure–Toxicity prediction of log(1/EC50) for some benzene derivatives from their density functional theory calculated molecular descriptors. *Bulletin of the Chemical Society of Japan*, *83*(4), 345–350. doi:10.1246/bcsj.20090309

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Hall, L. H., & Kier, L. B. (1995). Electrotopological state indices for atom types: A novel combination of electronic, topological, and valence state information. *Journal of Chemical Information and Computer Sciences*, *35*(6), 1039–1045. doi:10.1021/ci00028a014

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

Hannongbua, S., Nivasanond, K., Lawtrakul, L., Pungpo, P., & Wolschann, P. (2001). 3D-Quantitative structure−activity relationships of HEPT derivatives as HIV-1 reverse transcriptase inhibitors, based on ab initio calculations. *Journal of Chemical Information and Computer Sciences*, *41*(3), 848–855. doi:10.1021/ci0001278 PMID:11410067

Hansch, C. (1993). Quantitative structure-activity relationships and the unnamed science. *Accounts of Chemical Research*, *26*(4), 147–153. doi:10.1021/ar00028a003

Hansch, C., & Clayton, J. M. (1973). Lipophilic character and biological activity of drugs II: The parabolic case. *Journal of Pharmaceutical Sciences*, *62*(1), 1–21. doi:10.1002/jps.2600620102 PMID:4566530

Hansch, C., & Fujita, T. (1995). *Classical and three-dimensional QSAR in agrochemistry*. ACS Publications. doi:10.1021/bk-1995-0606

Hansch, C., & Gao, H. (1997). Comparative QSAR: Radical reactions of benzene derivatives in chemistry and biology. *Chemical Reviews*, *97*(8), 2995–3060. doi:10.1021/cr9601021 PMID:11851484

Hansch, C., Gao, H., & Hoekman, D. (1998). *A generalized approach to comparative QSAR*. Washington, DC: Taylor and Francis.

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., Kim, D., Leo, A. J., Novellino, E., Silipo, C., & Vittoria, A. et al. (1989). Toward a quantitative comparative toxicology of organic compounds. *CRC Critical Reviews in Toxicology*, *19*(3), 185–226. doi:10.3109/10408448909037471 PMID:2653732

Hansch, C., & Klein, T. E. (1986). Molecular graphics and QSAR in the study of enzyme-ligand interactions. On the definition of bioreceptors. *Accounts of Chemical Research*, *19*(12), 392–400. doi:10.1021/ar00132a003

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., Hoekman, D., & Heller, S. R. (1995). *Exploring Qsar*. Washington, DC: American Chemical Society.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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., & Steward, A. R. (1964). The use of substituent constants in the analysis of the structure–activity relationship in penicillin derivatives. *Journal of Medicinal Chemistry*, *7*(6), 691–694. doi:10.1021/jm00336a001 PMID:14262788

Hansen, C., Telzer, B. R., & Zhang, L. (1995). Comparative QSAR in toxicology: Examples from teratology and cancer chemotherapy of aniline mustards. *Critical Reviews in Toxicology*, *25*(1), 67–89. doi:10.3109/10408449509089887 PMID:7734060

Hemalatha, T., Imran, P. K. M., Gnanamani, A., & Nagarajan, S. (2012). QSAR and evaluation of molecular electrostatic potential for N-nitrosopiperidinone semicarbazones. *Chemometrics and Intelligent Laboratory Systems*, *116*, 87–93. doi:10.1016/j.chemolab.2012.04.014

Hilal, R., & Elroby, S. A. K. (2011). A QSAR study for 2-(4-aminophenyl)benzothiazoles: Using DFT optimisation of geometry of molecules. *Molecular Simulation*, *37*(1), 62–71. doi:10.1080/08927022.2010.520133

Hilal, S. H., Carreira, L. A., & Karickhoff, S. W. (1994). Estimation of chemical reactivity parameter and physical properties of organic molecules using SPARC. In P. Politzer & J. S. Murray (Eds.), *Quantitative treatments of solute/solvent interactions* (pp. 291–353). Amsterdam: Elsevier.

Hohenberg, P., & Kohn, W. (1964). Inhomogeneous electron gas. *Physical Review B: Condensed Matter and Materials Physics*, *136*(3B), B864–B871. doi:10.1103/PhysRev.136.B864

Hughes, M. F. (2002). Arsenic toxicity and potential mechanisms of action. *Toxicology Letters*, *133*(1), 1–16. doi:10.1016/S0378-4274(02)00084-X PMID:12076506

Jayatilleke, P. R. N., Nair, A. C., Zauhar, R., & Welsh, W. J. (2000). Computational studies on HIV-1 protease inhibitors: Influence of calculated inhibitor−enzyme binding affinities on the statistical quality of 3D-QSAR CoMFA models. *Journal of Medicinal Chemistry*, *43*(23), 4446–4451. doi:10.1021/jm9905357 PMID:11087569

Kamlet, M. J., Abboud, J.-L. M., Abraham, M. H., & Taft, R. W. (1983). Linear solvation energy relationships. 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. (2010). Predictive toxicology using QSAR: A perspective. *Journal of the Indian Chemical Society*, *87*, 1455–1515.

Karabunarliev, S., Mekenyan, O. G., Karcher, W., Russom, C. L., & Bradbury, S. P. (1996). Quantum-chemical descriptors for estimating the acute Ttoxicity of substituted benzenes to the Guppy (*Poecilia reticulata*) and Fathead Minnow (*Pimephales promelas*). *Quantitative Structure-Activity Relationships*, *15*(4), 311–320. doi:10.1002/qsar.19960150406

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Karcher, W., & Devillers, J. (1990). Practical applications of quantitative structure-activity relationships (QSAR). *Environmental Chemistry and Toxicology, 1*.

Karelson, M., Lobanov, V. S., & Katritzky, A. R. (1996). Quantum-chemical descriptors in QSAR/QSPR studies. *Chemical Reviews*, *96*(3), 1027–1043. doi:10.1021/cr950202r PMID:11848779

Katritzky, A. R., Fara, D. C., Petrukhin, R. O., Tatham, D. B., Maran, U., Lomaka, A., & Karelson, M. (2002). The present utility and future potential for medicinal chemistry of QSAR / QSPR with whole molecule descriptors. *Current Topics in Medicinal Chemistry*, *2*(12), 1333–1356. doi:10.2174/1568026023392922 PMID:12470284

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., Maran, U., Lobanov, V. S., & Karelson, M. (2000). Structurally diverse quantitative structure-property relationship correlations of technologically relevant physical properties. *Journal of Chemical Information and Computer Sciences*, *40*(1), 1–18. doi:10.1021/ci9903206 PMID:10661545

Katritzky, A. R., Mu, L., Lobanov, V. S., & Karelson, M. (1996). Correlation of boiling points with molecular structure. A training set of 298 diverse organics and a test set of 9 simple inorganics. *Journal of Physical Chemistry*, *100*(24), 10400–10407. doi:10.1021/jp953224q

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

Kim, K. H. (1995). Molecular similarity. In P. M. Dean (Ed.), *Drug design* (pp. 291–331). New York: Chapman & Hall; doi:10.1007/978-94-011-1350-2_12

Kochakian, C. D. (1976). *Anabolic-androgenic steroids*. Berlin: Springer. doi:10.1007/978-3-642-66353-6

Kubinyi, H. (Ed.). (1993). *3D QSAR in drug design: Theory, methods and applications*. Leiden, The Netherlands: ESCOM.

Kubinyi, H. (1997). QSAR and 3D QSAR in drug design part 2: Applications and problems. *Drug Discovery Today*, *2*(12), 538–546. doi:10.1016/S1359-6446(97)01084-2

Kubinyi, H. (1998). Comparative molecular field analysis (CoMFA). In N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer III, & P. R. Schreiner (Eds.), *The encyclopedia of computational chemistry* (pp. 448–460). Chichester, UK: Wiley.

Kumar, D., Singh, R. N., Sahu, S., & Baboo, V. (2011). Hardness based quantitative structure toxicity relationship (QSTR) study on a series of aliphatic alcohol derivatives. *Organic Chemistry: An Indian Journal*, *7*, 41–47.

Langridge, R., & Klein, T. (1990). Molecular graphics and drug design. Comprehensive Medicinal Chemistry, 4, 413-429.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Leermakers, M., Baeyens, W., De Gieter, M., Smedts, B., Meert, C., & De Bisschop, H. C. et al. (2006). Toxic arsenic compounds in environmental samples: Speciation and validation. *Trends in Analytical Chemistry*, *25*(1), 1–10. doi:10.1016/j.trac.2005.06.004

Leo, A. J., & Hansch, C. (1999). Role of hydrophobic effects in mechanistic QSAR. *Perspectives in Drug Discovery and Design*, *17*(1), 1–25. doi:10.1023/A:1008762321231

Levet, A., Bordes, C., Clément, Y., Mignon, P., Chermette, H., & Marote, P. et al. (2013). Quantitative structure-activity relationship to predict acute fish toxicity of organic solvents. *Chemosphere*, *93*, 1094–1103. PMID:23866172

Li, Z., Ning, C., Bin, Y., GuangFu, Y., Yun, L., & XinLing, Y. (2010). QSAR and 3D-QSAR studies of the diacyl-hydrazine derivatives containing furan rings based on the density functional theory. *SCIENCE CHINA Chemistry, 53*, 1322-1331.

Liao, S., Liang, T., Fang, S., Castenada, E., & Shao, T. C. (1973). Steroid structure and androgenic activity specificities involved in the receptor binding and nuclear retention of various androgens. *The Journal of Biological Chemistry*, *248*, 6154–6162. PMID:4353633

Lien, E. J., Hansch, C., & Anderson, S. M. (1968). Structure-activity correlations for antibacterial agents on gram-positive and gram-negative cells. *Journal of Medicinal Chemistry*, *11*(3), 430–441. doi:10.1021/jm00309a004 PMID:5656475

Lucic, B., & Trinajstic, N. (1999). Multivariate regression outperforms several robust architectures of neural networks in QSAR modeling. *Journal of Chemical Information and Computer Sciences*, *39*(1), 121–132. doi:10.1021/ci980090f

Maddalena, D. J. (1998). Applications of soft computing in drug design. *Expert Opinion on Therapeutic Patents*, *8*(3), 249–258. doi:10.1517/13543776.8.3.249

Martin, Y. C., Kim, K.-H., & Lin, C. T. (1996). Article. *Advances in Quantitative Structure Property Relationships*, *1*, 1–52.

Martin, Y. C., & Lin, C. T. (1996). Article. In The practice of medicinal chemistry. London: Academic Press.

Matta, C. F. (2014). Modeling biophysical and biological properties from the characteristics of the molecular electron density, electron localization and delocalization matrices, and the electrostatic potential. *Journal of Computational Chemistry*, *35*(16), 1165–1198. doi:10.1002/jcc.23608 PMID:24777743

Maynard, A. T., Huang, M., Rice, W. G., & Covel, D. G. (1998). Reactivity of the HIV-1 nucleocapsid protein p7 zinc finger domains from the perspective of density-functional theory. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(20), 11578–11583. doi:10.1073/pnas.95.20.11578 PMID:9751708

McFarland, J. W. (1970). Parabolic relation between drug potency and hydrophobicity. *Journal of Medicinal Chemistry*, *13*(6), 1092–1196. doi:10.1021/jm00300a040 PMID:5479848

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Mezey, P. G. (1992). Shape-similarity measures for molecular bodies: A 3D topological approach to quantitative shape-activity relations. *Journal of Chemical Information and Computer Sciences*, *32*(6), 650–656. doi:10.1021/ci00010a011

Mishra, S. K., Mishra, V., Tripathi, P. N., & Khan, M. A. (2014). Quantum chemical descriptors based QSTR study of nitrobenzene derivatives against tetrahymena pyriformis. *Research Journal of Chemical Sciences*, *4*, 29–37.

Mo, L.-Y., Liu, H.-Y., Yi, Z.-S., Li, Y.-H., & Dou, R.-N. (2011). QSTR study on the freshwater photobacteria toxicity of substituted benzenes. *Chinese Journal of Structural Chemistry*, *30*, 1403–1411.

Moudgal, C. J., Lipscomb, J. C., & Bruce, R. M. (2000). Potential health effects of drinking water disinfection by-products using quantitative structure toxicity relationship. *Toxicology*, *147*(2), 109–131. doi:10.1016/S0300-483X(00)00188-8 PMID:10874158

Mulliken, R. S. (1934). A new electroaffinity scale; together with data on valence states and on valence ionization potentials and electron affinities. *The Journal of Chemical Physics*, *2*(11), 782–793. doi:10.1063/1.1749394

Mulliken, R. S. (1935). Electronic structures of molecules XI. Electroaffinity, molecular orbitals and dipole moments. *The Journal of Chemical Physics*, *3*(9), 573–585. doi:10.1063/1.1749731

Murray, J. S., & Politzer, P. (2003). The use of the molecular electrostatic potential in medicinal chemistry. In P. Carloni & F. Alber (Eds.), *Quantum medicinal chemistry*. Weinheim: Wiley-VCH.

Murray, J. S., & Politzer, P. A. (1994). General interaction properties function (GIPF): an approach to understanding and predicting molecular interactions. In P. Politzer & J. S. Murray (Eds.), *Quantitative treatments of solute/solvent interactions* (pp. 243–289). Amsterdam: Elsevier.

Nakajoh, K., Shibata, E., Todoroki, T., Ohara, A., Nishizawa, K., & Nakamura, T. (2006). Measurement of temperature dependence for the vapor pressures of twenty six polychlorinated biphenyl congeners in commercial Kanechlor mixtures by the Knudsen effusion method. *Environmental Toxicology and Chemistry*, *25*(2), 327–336. doi:10.1897/05-215R.1 PMID:16519292

Nantasenamat, C., Isarankura-Na-Ayudhya, C., Naenna, T., & Prachayasittikul, V. (2009). A practical overview of quantitative structure-activity relationship. *Excli Journal.*, *8*, 74–88.

Norinder, U. (1998). Recent progress in CoMFA methodology and related techniques. *Perspectives in Drug Discovery and Design*, *12*, 25–39. doi:10.1023/A:1017057332537

O'Brien, S. E., & Popelier, P. L. A. (1999). Quantum molecular similarity. Part 2: The relation between properties in BCP space and bond length. *Canadian Journal of Chemistry*, *77*(1), 28–36. doi:10.1139/v98-215

Padmanabhan, J., Parthasarathi, R., Subramanian, V., & Chattaraj, P. K. (2006a). Theoretical study on the complete series of chloroanilines. *The Journal of Physical Chemistry A*, *110*(32), 9900–9907. doi:10.1021/jp061436p PMID:16898692

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Padmanabhan, J., Parthasarathi, R., Subramanian, V., & Chattaraj, P. K. (2006b). QSPR models for polychlorinated biphenyls: *N*-Octanol/water partition coefficient. *Bioorganic & Medicinal Chemistry*, *14*(4), 1021–1028. doi:10.1016/j.bmc.2005.09.017 PMID:16214354

Padmanabhan, J., Parthasarathi, R., Subramanian, V., & Chattaraj, P. K. (2006c). Group philicity and electrophilicity as possible descriptors for modeling ecotoxicity applied to chlorophenols. *Chemical Research in Toxicology*, *19*(3), 356–364. doi:10.1021/tx050322m PMID:16544939

Padmanabhan, J., Parthasarathi, R., Subramanian, V., & Chattaraj, P. K. (2007). Using QSPR models to predict the enthalpy of vaporization of 209 polychlorinated biphenyl congeners. *QSAR & Combinatorial Science*, *26*(2), 227–237. doi:10.1002/qsar.200630027

Pal, S., & Chandra, A. K. (1995). Some novel relationships of polarizability with dipole moments. *Journal of Physical Chemistry*, *99*(38), 13865–13867. doi:10.1021/j100038a017

Pan, S., & Chattaraj, P. K. (2013). Favorable direction in a chemical reaction through the maximum hardness principle. *Journal of the Mexican Chemical Society*, *57*, 23–24.

Pan, S., Solà, M., & Chattaraj, P. K. (2013). On the validity of the maximum hardness principle and the minimum electrophilicity principle during chemical reactions. *The Journal of Physical Chemistry A*, *117*(8), 1843–1852. doi:10.1021/jp312750n PMID:23373511

Pandith, A. H., & Islam, N. (2013). Comparative assessment of QSTR models based on density functional, hartree–fock, AM1, and PM3 methods for acute toxicity of aliphatic compounds toward Vibrio fischeri. *International Journal of Quantum Chemistry*, *113*(6), 830–839. doi:10.1002/qua.24133

Parr, R. G. (1922-1924). Szentpaly, Lv., & Liu, S. (1999). Electrophilicity index. *Journal of the American Chemical Society*, 121.

Parr, R. G., & Chattaraj, P. K. (1991). Principle of maximum hardness. *Journal of the American Chemical Society*, *113*(5), 1854–1855. doi:10.1021/ja00005a072

Parr, R. G., Donnelly, R. A., Levy, M., & Palke, W. E. (1978). Electronegativity: The density functional viewpoint. *The Journal of Chemical Physics*, *68*(8), 3801–3807. doi:10.1063/1.436185

Parr, R. G., & Pearson, R. G. (1983). Absolute hardness: Companion parameter to absolute electronegativity. *Journal of the American Chemical Society*, *105*(26), 7512–7516. doi:10.1021/ja00364a005

Parr, R. G., & Yang, W. (1984). Density functional approach to the frontier-electron theory of chemical reactivity. *Journal of the American Chemical Society*, *106*(14), 4049–4050. doi:10.1021/ja00326a036

Parr, R. G., & Yang, W. (1989). *Density functional theory of atoms and molecules*. New York: Oxford University Press.

Parr, R. G., & Yang, W. (1995). Density-functional theory of the electronic structure of molecules. *Annual Review of Physical Chemistry*, *46*(1), 701–728. doi:10.1146/annurev.pc.46.100195.003413 PMID:24341393

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Parr, R. G., & Zhou, Z. (1993). Absolute hardness: Unifying concept for identifying shells and subshells in nuclei, atoms, molecules, and metallic clusters. *Accounts of Chemical Research*, *26*(5), 256–258. doi:10.1021/ar00029a005

Parthasarathi, R., Elango, M., Subramanian, V., & Chattaraj, P. K. (2005). Variation of electrophilicity during molecular vibrations and internal rotations. *Theoretical Chemistry Accounts*, *113*(5), 257–266. doi:10.1007/s00214-005-0634-3

Parthasarathi, R., Padmanabhan, J., Elango, M., Subramanian, V., & Chattaraj, P. K. (2004a). Intermolecular reactivity through the generalized philicity concept. *Chemical Physics Letters*, *394*(4-6), 225–230. doi:10.1016/j.cplett.2004.07.002

Parthasarathi, R., Subramanian, V., Roy, D. R., & Chattaraj, P. K. (2004b). Electrophilicity index as a possible descriptor of biological activity. *Bioorganic & Medicinal Chemistry*, *12*(21), 5533–5543. doi:10.1016/j.bmc.2004.08.013 PMID:15465330

Pasha, F. A., Srivastava, H. K., & Singh, P. P. (2005). QSAR study of estrogens with the help of PM3-based descriptors. *International Journal of Quantum Chemistry*, *104*(1), 87–100. doi:10.1002/qua.20569

Pasha, F. A., Srivastava, H. K., & Singh, P. P. (2005). Semiempirical QSAR study and ligand receptor interaction of estrogens. *Molecular Diversity*, *9*(1-3), 215–220. doi:10.1007/s11030-005-2711-x PMID:15789567

Pathak, R. K., & Gadre, S. R. (1990). Maximal and minimal characteristics of molecular electrostatic potentials. *The Journal of Chemical Physics*, *93*(3), 1770–1773. doi:10.1063/1.459703

Patil, G. S. (1991). Correlation of aqueous solubility and octanol-water partition coefficient based on molecular structure. *Chemosphere*, *22*(8), 723–738. doi:10.1016/0045-6535(91)90049-J

Pauling, L. (1932). The nature of the chemical bond. Iv. The energy of single bonds and the relative electronegativity of atoms. *Journal of the American Chemical Society*, *54*(9), 3570–3582. doi:10.1021/ja01348a011

Pauling, L. (1960). *The nature of the chemical bond*. Ithaca, NY: Cornell University Press.

Pearson, R. G. (1963). Hard and soft acids and bases. *Journal of the American Chemical Society*, *85*(22), 3533–3539. doi:10.1021/ja00905a001

Pearson, R. G. (1966). Acids and bases. *Science*, *151*(3707), 172–177. doi:10.1126/science.151.3707.172 PMID:17746330

Pearson, R. G. (1968a). Hard and soft acids and bases, HSAB, part 1: Fundamental principles. *Journal of Chemical Education*, *45*(9), 581–587. doi:10.1021/ed045p581

Pearson, R. G. (1968b). Hard and soft acids and bases, HSAB, part II: Underlying theories. *Journal of Chemical Education*, *45*(10), 643–648. doi:10.1021/ed045p643

Pearson, R. G. (1986). Absolute electronegativity and hardness correlated with molecular orbital theory. *Proceedings of the National Academy of Sciences of the United States of America*, *83*(22), 8440–8441. doi:10.1073/pnas.83.22.8440 PMID:16578791

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Pearson, R. G. (1987). Recent advances in the concept of hard and soft acids and bases. *Journal of Chemical Education*, *64*(7), 561–567. doi:10.1021/ed064p561

Pearson, R. G. (1993). The principle of maximum hardness. *Accounts of Chemical Research*, *26*(5), 250–255. doi:10.1021/ar00029a004

Pearson, R. G. (1997). *Chemical hardness: Applications from molecules to solids*. Weinheim: Wiley-VCH. doi:10.1002/3527606173

Politzer, P. (1987). A relationship between the charge capacity and the hardness of neutral atoms and groups. *The Journal of Chemical Physics*, *86*(2), 1072–1073. doi:10.1063/1.452296

Politzer, P. (1988). Computational approaches to the identification of suspect toxic molecules. *Toxicology Letters*, *43*(1-3), 257–276. doi:10.1016/0378-4274(88)90032-X PMID:3051523

Popelier, P. L. A. (1999). Quantum molecular similarity. 1. BCP space. *The Journal of Physical Chemistry A*, *103*(15), 2883–2890. doi:10.1021/jp984735q

Popelier, P. L. A., Smith, P. J., & Chaudry, U. A. (2004). Quantitative structure-activity relationships of mutagenic activity from quantum topological descriptors: Triazenes and halogenated hydroxyfuranones (mutagen-X) derivatives. *Journal of Computer-Aided Molecular Design*, *18*(11), 709–718. doi:10.1007/s10822-004-6815-7 PMID:15865063

Puri, S., Chickos, J. S., & Welsh, W. J. (2001). Determination of vaporization enthalpies of polychlorinated biphenyls by correlation gas chromatography. *Analytical Chemistry*, *73*(7), 1480–1484. doi:10.1021/ac001246p PMID:11321297

Puzyn, T., Leszczyński, J., & Cronin, M. T. D. (2010). *Recent advances in QSAR studies: Methods and applications* (Vol. 8). Springer. doi:10.1007/978-1-4020-9783-6

Randić, M. (1975). Characterization of molecular branching. *Journal of the American Chemical Society*, *97*(23), 6609–6615. doi:10.1021/ja00856a001

Randić, M. (1991). Generalized molecular descriptors. *Journal of Mathematical Chemistry*, *7*(1), 155–168. doi:10.1007/BF01200821

Randić, M., & Razinger, M. (1996). On the characterization of three-dimensional molecular structure. In From chemical topology to three-dimensional. Plenum Press.

Rasulev, B. F., Saidkhodzhaev, A. I., Nazrullaev, S. S., Akhmedkhodzhaeva, K. S., Khushbaktova, Z. A., & Leszczynski, J. (2007). Molecular modelling and QSAR analysis of the estrogenic activity of terpenoids isolated from Ferula plants. *SAR and QSAR in Environmental Research*, *18*(7-8), 663–673. doi:10.1080/10629360701428631 PMID:18038366

Rekker, R. F., & Mannhold, R. (1992). *Calculation of drug lipophilicity*. Weinheim: VCH.

Rohrbaugh, R. H., & Jurs, P. C. (1987). Descriptions of molecular shape applied in studies of structure/activity and structure/property relationships. *Analytica Chimica Acta*, *199*, 99–109. doi:10.1016/S0003-2670(00)82801-9

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Rokhina, E. V., & Suri, R. P. S. (2012). Application of density functional theory (DFT) to study the properties and degradation of natural estrogen hormones with chemical oxidizers. *The Science of the Total Environment*, *417–418*, 280–290. doi:10.1016/j.scitotenv.2011.12.008 PMID:22245212

Roy, D. R., Giri, S., & Chattaraj, P. K. (2009). Arsenic toxicity: An atom counting and electrophilicity-based protocol. *Molecular Diversity*, *13*(4), 551–556. doi:10.1007/s11030-009-9133-0 PMID:19326252

Roy, D. R., Pal, N., Mitra, A., Bultinck, P., Parthasarathi, R., Subramanian, V., & Chattaraj, P. K. (2007). An atom counting strategy towards analyzing the biological activity of sex hormones. *European Journal of Medicinal Chemistry*, *42*(11-12), 1365–1369. doi:10.1016/j.ejmech.2007.01.028 PMID:17416442

Roy, D. R., Parthasarathi, R., Padmanabhan, J., Sarkar, U., Subramanian, V., & Chattaraj, P. K. (2006a). Careful scrutiny of the philicity concept. *The Journal of Physical Chemistry A*, *110*(3), 1084–1093. doi:10.1021/jp053641v PMID:16420012

Roy, D. R., Parthasarathi, R., Subramanian, V., & Chattaraj, P. K. (2006b). An electrophilicity based analysis of toxicity of aromatic compounds towards Tetrahymena pyriformis. *QSAR & Combinatorial Science*, *25*(2), 114–122. doi:10.1002/qsar.200530146

Roy, D. R., Sarkar, U., Chattaraj, P. K., Mitra, A., Padmanabhan, J., & Parthasarathi, R. et al. (2006c). Analyzing toxicity through electrophilicity. *Molecular Diversity*, *10*(2), 119–131. doi:10.1007/s11030-005-9009-x PMID:16763875

Roy, K., & Ghosh, G. (2004). QSTR with extended topochemical atom indices. 3. Toxicity of nitrobenzenes to tetrahymena pyriformis. *QSAR & Combinatorial Science*, *23*(23), 99–108. doi:10.1002/qsar.200330864

Roy, K., & Mitra, I. (2009). Advances in quantitative structure-activity relationship models of antioxidants. *Expert Opinion on Drug Discovery*, *4*(11), 1157–1175. doi:10.1517/17460440903307409 PMID:23480434

Russom, C. L., Bradbury, S. P., Broderius, S. J., Hammermeister, D. E., & Drummond, R. A. (1997). Predicting modes of toxic action from chemical structure: Acute toxicity in the fathead minnow (*Pimephales promelas*). *Environmental Toxicology and Chemistry*, *16*(5), 948–967. doi:10.1897/1551-5028(1997)016<0948:PMOTAF>2.3.CO;2

Saartok, T., Dahlberg, E., & Gustafsson, J. A. (1984). Relative binding affinity of anabolic-androgenic steroids: Comparison of the binding to the androgen receptors in skeletal muscle and in prostate, as well as to sex hormone-binding globulin. *Endocrinology*, *114*(6), 2100–2106. doi:10.1210/endo-114-6-2100 PMID:6539197

Sala, G., & Baldratti, G. (1957). Myotrophic (anabolic) activity of 4-substituted testosterone analogs. *Proceedings of the Society for Experimental Biology and Medicine New York*, *95*, 22-24. doi:10.3181/00379727-95-23105

Schultz, T. W. (1997). Tetratox: Tetrahymena pyriformis population growth impairment endpointa surrogate for fish lethality. *Toxicology Mechanisms and Methods*, *7*(4), 289–309. doi:10.1080/105172397243079

Schultz, T. W. (1999). Structure−toxicity relationships for benzenes evaluated with *Tetrahymena pyriformis. Chemical Research in Toxicology, 12*, 1262-1267.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Schultz, T. W., Cronin, M. T. D., & Netzeva, T. I. (2003a). The present status of QSAR in toxicology. *Journal of Molecular Structure THEOCHEM*, *622*(1-2), 23–38. doi:10.1016/S0166-1280(02)00615-2

Schultz, T. W., Cronin, M. T. D., Netzeva, T. I., & Aptula, A. O. (2002). Structure−toxicity relationships for aliphatic chemicals evaluated with *Tetrahymena pyriformis. Chemical Research in Toxicology*, *15*(12), 1602–1609. doi:10.1021/tx025589p PMID:12482243

Schultz, T. W., Cronin, M. T. D., Walker, J. D., & Aptul, A. O. (2003b). Quantitative structure–activity relationships (QSARs) in toxicology: A historical perspective. *Journal of Molecular Structure THEOCHEM*, *622*(1-2), 1–22. doi:10.1016/S0166-1280(02)00614-0

Schultz, T. W., Netzeva, T. I., & Cronin, M. T. D. (2003c). Selection of data sets for QSARs: Analyses of Tetrahymena toxicity from aromatic compounds. *SAR and QSAR in Environmental Research*, *14*(1), 59–81. doi:10.1080/1062936021000058782 PMID:12688416

Schuur, J., Selzer, P., & Gasteiger, J. (1996). The coding of the three-dimensional structure of molecules by molecular transforms and its application to structure-spectra correlations and studies of biological activity. *Journal of Chemical Information and Computer Sciences*, *36*(2), 334–344. doi:10.1021/ci950164c

Selassie, C., Mekapati, S., & Verma, R. (2002). QSAR: Then and now. *Current Topics in Medicinal Chemistry*, *2*(12), 1357–1379. doi:10.2174/1568026023392823 PMID:12470285

Selassie, C., & Verma, R. P. (2010). History of quantitative structure–activity relationships. In *Burger's medicinal chemistry and drug discovery*. Academic Press.

Senior, S. A., Madbouly, M. D., & El massry, A.-M. (2011). QSTR of the toxicity of some organophosphorus compounds by using the quantum chemical and topological descriptors. *Chemosphere*, *85*(1), 7–12. doi:10.1016/j.chemosphere.2011.06.062 PMID:21757222

Simon-Manso, Y., & Fuentealba, P. (1998). On the density functional relationship between static dipole polarizability and global softness. *The Journal of Physical Chemistry A*, *102*(11), 2029–2032. doi:10.1021/jp972733k

Singh, P. P., Srivastava, H. K., & Pasha, F. A. (2004). DFT-based QSAR study of testosterone and its derivatives. *Bioorganic & Medicinal Chemistry*, *12*(1), 171–177. doi:10.1016/j.bmc.2003.11.002 PMID:14697782

Singhal, K., Sahu, V. K., Singh, P., & Raj, P. (2014). DFT-based prediction of antifungal and insecticidal activities of perfluorophenyl antimony(III) and antimony(V) chlorides. *Medicinal Chemistry Research*, *23*(4), 1758–1767. doi:10.1007/s00044-013-0752-8

Sippl, W., & Holtje, H. D. (2000). Structure-based 3D-QSAR—Merging the accuracy of structure-based alignments with the computational efficiency of ligand-based methods. *Journal of Molecular Structure THEOCHEM*, *503*(1-2), 31–50. doi:10.1016/S0166-1280(99)00361-9

Smith, P. J., & Popelier, P. L. A. (2005). Quantum chemical topology (QCT) descriptors as substitutes for appropriate Hammett constants. *Organic & Biomolecular Chemistry*, *3*(18), 3399–3407. doi:10.1039/b507024d PMID:16132102

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Song, J., Wang, Z., Findlater, A., Han, Z., Jiang, Z., & Chen, J. et al. (2013). Terpenoid mosquito repellents: A combined DFT and QSAR study. *Bioorganic & Medicinal Chemistry Letters*, *23*(5), 1245–1248. doi:10.1016/j.bmcl.2013.01.015 PMID:23375229

Srivastava, H. K., Pasha, F. A., & Singh, P. P. (2005). Atomic softness-based QSAR study of testosterone. *International Journal of Quantum Chemistry*, *103*(3), 237–245. doi:10.1002/qua.20506

Stanton, D. T., & Jurs, P. C. (1990). Development and use of charged partial surface area structural descriptors in computer-assisted quantitative structure–property relationship studies. *Analytical Chemistry*, *62*(21), 2323–2329. doi:10.1021/ac00220a013

Stuper, A. J., Brugger, W. E., & Jurs, P. C. (1979). *Computer assisted studies of chemical structure and biological function*. New York: Wiley-Interscience.

Taft, R. W. (1956). *Steric effects in organic chemistry* (M. S. Newman, Ed.). New York: John Wiley and Sons.

Taft, R. W., Abboud, J.-L. M., Kamlet, M. J., & Abraham, M. H. (1985). Linear solvation energy relations. *Journal of Solution Chemistry*, *14*(3), 153–186. doi:10.1007/BF00647061

Tanwar, A., Pal, S., Roy, D. R., & Chattaraj, P. K. (2006). Minimum magnetizability principle. *The Journal of Chemical Physics*, *125*(5), 056101–056102. doi:10.1063/1.2227381 PMID:16942262

Tawari, N., Lele, A., Khambete, M., & Degani, M. (2014). Mutagenicity prediction for nitroaromatic compounds using QSTR modeling. *International Journal of Pharmacy and Pharmaceutical Sciences*, *6*, 149–157.

Todeschini, R., & Consonni, V. (2000). *Hand book of molecular descriptors*. Weinheim: Wiley-VCH. doi:10.1002/9783527613106

Todeschini, R., Lasagni, M., & Marengo, E. (1994). New molecular descriptors for 2D- and 3D-structures. Theory. *Journal of Chemometrics*, *8*(4), 263–273. doi:10.1002/cem.1180080405

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

Tripathi, P. N., & Mishra, V. (2014). Toxicological study of nitrobenzene derivatives against tetrahymena pyriformis using topological parameters. *Journal of Chemical and Pharmaceutical Research*, *6*, 848–854.

Trohalaki, S., & Pachter, R. (2003). Quantum descriptors for predictive toxicology of halogenated aliphatic hydrocarbons. *SAR and QSAR in Environmental Research*, *14*(2), 131–143. doi:10.1080/1062936031000073153 PMID:12747571

van Vleck, J. (1932). *The theory of electric and magnetic susceptibilities*. Oxford, UK: Oxford University Press.

Veith, G. D., & Mekenyan, O. G. (1993). A QSAR approach for estimating the aquatic toxicity of soft electrophiles. *Quantitative Structure-Activity Relationships*, *12*(4), 349–356. doi:10.1002/qsar.19930120402

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Vela, A., & Gázquez, J. L. (1990). A relationship between the static dipole polarizability, the global softness, and the Fukui function. *Journal of the American Chemical Society*, *112*(4), 1490–1492. doi:10.1021/ja00160a029

Vijayaraj, R., Subramanian, V., & Chattaraj, P. K. (2009). Comparison of global reactivity descriptors calculated using various density functionals: A QSAR perspective. *Journal of Chemical Theory and Computation*, *5*(10), 2744–2753. doi:10.1021/ct900347f

Voskresensky, O. N., & Levitsky, A. P. (2002). QSAR aspects of flavonoids as a plentiful source of new drugs. *Current Medicinal Chemistry*, *9*(14), 1367–1383. doi:10.2174/0929867023369790 PMID:12132993

Wan, J., Zhang, L., Yang, G., & Zhan, C.-G. (2004). Quantitative structure-activity relationship for cyclic imide derivatives of protoporphyrinogen oxidase inhibitors: A study of quantum chemical descriptors from density functional theory. *Journal of Chemical Information and Computer Sciences*, *44*(6), 2099–2105. doi:10.1021/ci049793p PMID:15554680

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

Wolff, M. E. (1955). Therapeutic agents. In *Burger's medicinal chemistry and drug discovery* (Vol. 4). New York: John Wiley & Sons.

Yan, D., Jiang, X., Xu, S., Wang, L., Bian, Y., & Yu, G. (2008). Quantitative structure-toxicity relationship study of lethal concentration to tadpole (Bufo vulgaris formosus) for organophosphorous pesticides. *Chemosphere*, *71*(10), 1809–1815. doi:10.1016/j.chemosphere.2008.02.033 PMID:18395243

Yang, G.-F., & Huang, X. (2006). Development of quantitative structure-activity relationships and its application in rational drug design. *Current Pharmaceutical Design*, *12*(35), 4601–4611. doi:10.2174/138161206779010431 PMID:17168765

Yang, W., & Mortier, W. J. (1986). The use of global and local molecular parameters for the analysis of the gas-phase basicity of amines. *Journal of the American Chemical Society*, *108*(19), 5708–5711. doi:10.1021/ja00279a008 PMID:22175316

Yang, W., & Parr, R. G. (1985). Hardness, softness, and the fukui function in the electronic theory of metals and catalysis. *Proceedings of the National Academy of Sciences of the United States of America*, *82*(20), 6723–6726. doi:10.1073/pnas.82.20.6723 PMID:3863123

Yang, W., Parr, R. G., & Pucci, R. (1984). Electron density, Kohn– Sham frontier orbitals, and Fukui functions. *The Journal of Chemical Physics*, *81*(6), 2862–2863. doi:10.1063/1.447964

Zhang, L., Hao, G.-F., Tan, Y., Xi, Z., Huang, M.-Z., & Yang, G.-F. (2009). Bioactive conformation analysis of cyclic imides as protoporphyrinogen oxidase inhibitor by combining DFT calculations, QSAR and molecular dynamic simulations. *Bioorganic & Medicinal Chemistry*, *17*(14), 4935–4942. doi:10.1016/j.bmc.2009.06.003 PMID:19540767

Zhang, S. G., Lei, W., Xia, M. Z., & Wang, F. Y. (2005). QSAR study on N-containing corrosion inhibitors: Quantum chemical approach assisted by topological index. *Journal of Molecular Structure THEOCHEM*, *732*(1-3), 173–182. doi:10.1016/j.theochem.2005.02.091

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

KEY TERMS AND DEFINITIONS

Aliphatic Compounds: These are non-aromatic compounds made by carbon and hydrogen.

Electrophilicity: It is a measure of 'electrophilic power' of a molecule. It can be estimated as the electronegativity squared divided by twice the hardness.

Energies of Frontier Molecular Orbitals: They are the energies of the highest occupied molecular orbital and lowest unoccupied molecular orbital, respectively.

Group Philicity: It is a condensed philicity summed over a group of relevant atoms.

Net Atomic Charge: It is a non-integer charge at each atom in a molecule in elementary charge units.

Net Electrophilicity: It is the electron-accepting power of a molecule relative to its own electron-donating power.

Polyaromatic Hydrocarbons: These are compounds made by only carbon and hydrogen and contain multiple aromatic rings.

Polychlorinated Biphenyls: These are compounds containing two phenyl rings and several chlorine atoms are attached with phenyl rings.

Tetrahymena Pyriformis: These are free-living ciliate protozoa, generally found in freashwater ponds.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

APPENDIX: LIST OF SYMBOLS AND ABBREVIATIONS

SAR: Structure-Activity Relationship
QSAR: Quantitative Structure-Activity Relationship
QSTR: Quantitative Structure-Toxicity Relationship
DFT: Density Functional Theory
CDFT: Conceptual Density Functional Theory
LSFER: Linear Solvation Free Energy Relationship
CoMFA: Comparative Molecular Field Analysis
QSPR: Quantitative Structure-Property Relationship
PCA: Principal Component Analysis
η**:** Hardness
χ**:** Electronegativity
ω**:** Electrophilicity
$f(\mathbf{r})$ **:** Fukui function
f_k**:** Condensed Fukui function
$s(\mathbf{r})$**:** Local softness
$\eta(\mathbf{r})$**:** Local hardness
$\omega(\mathbf{r})$**:** Philicity
IP: Ionization Potential
EA: Electron Affinity
μ**:** Chemical Potential
E: Energy
N: Number of Electrons
$v(\mathbf{r})$**:** External Potential
$\Delta\omega^{\pm}$**:** Net Electrophilicity
HIV-1: Human Immunodficiency Virus type 1
NCp7: Nucleocapsid Protein p7
W: Electrical Power
V: Voltage
R: Electrical Resistance
α**:** Polarizability
ξ**:** Magnetizability
$\rho(\mathbf{r})$**:** Electron Density
q_k**:** Atomic Charge at k^{th} site
ω_g^{α}**:** Group Philicity
PMH: Principle of Maximum Hardness
MEP: Minimum Electrophilicity Principle
MPP: Minimum Polarizability Principle
MMP: Minimum Magnetizability Principle
$\mathbf{LD_{50}}$**:** 50% Lethal Dose
$\mathbf{LC_{50}}$**:** 50% Lethal Concentration

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Z: Atomic Number
N_{NH}**:** Number of Nonhydrogenic Atoms
R or r: Coefficient of correlation
$\mathbf{R^2_{CV}}$ **or** $\mathbf{r^2_{CV}}$**:** Variance of Leave-one-out Cross-Validation
PAH: Polyaromatic Hydrocarbons
TCDD: Tetrachlorodibenzo-p-Dioxin
Ah: Arylhydrocarbon
PCDF: Polychlorinated Dibenzofurans
CA: Chloroanilines
PCB: Polychlorinated Biphenyl
RBA: Relative Binding Affinity
TeBG: Testosterone-Binding Globulins
MAPT Myotrophic to Androgenic Potency in Temporal Activity.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 5
Importance of Applicability Domain of QSAR Models

Kunal Roy
Jadavpur University, India

Supratik Kar
Jadavpur University, India

ABSTRACT

Quantitative Structure-Activity Relationship (QSAR) models have manifold applications in drug discovery, environmental fate modeling, risk assessment, and property prediction of chemicals and pharmaceuticals. One of the principles recommended by the Organization of Economic Co-operation and Development (OECD) for model validation requires defining the Applicability Domain (AD) for QSAR models, which allows one to estimate the uncertainty in the prediction of a compound based on how similar it is to the training compounds, which are used in the model development. The AD is a significant tool to build a reliable QSAR model, which is generally limited in use to query chemicals structurally similar to the training compounds. Thus, characterization of interpolation space is significant in defining the AD. An attempt is made in this chapter to address the important concepts and methodology of the AD as well as criteria for estimating AD through training set interpolation in the descriptor space.

INTRODUCTION

Quantitative structure-activity relationship (QSAR) modelling has become an important tool in the new drug candidate design, environmental fate modeling, toxicity and property prediction of chemicals and pharmaceuticals since they offer an economical and time-effective alternative to the medium throughput *in vitro* and low throughput *in vivo* assays (Perkins *et al*., 2003; Selassie, 2003; Walker et al., 2003). A QSAR model is a simple mathematical equation that is evaluated from a set of molecules with known activities/properties/toxicities using computational approaches. Thereafter, the developed predictive QSAR models are also applied by regulatory agencies to evaluate physical, chemical, and biological properties of individual chemical entities using applications specific for decision-making frameworks in risk and safety assessments (Kar & Roy, 2010). The QSAR modeling hypothesis also supports the 3Rs

DOI: 10.4018/978-1-4666-8136-1.ch005

Copyright © 2015, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

(replacement, refinement and reduction in animals in research) paradigm due to an increased pressure from social and economic background to trim down the use of animal testing as an important alternative method for future prediction of untested chemical entities (Benigni & Giuliani, 2003).

Robust validation of QSAR models plays a key step for the selection of a predictive model that may be considered for future prediction of new molecules. As numerous numbers of researches have been directed to the design of new molecules with the utilization of QSAR technique, validation of a QSAR model has been certified as the most considerable stride (Carlsen *et al*., 2009) for assessing the quality of data, applicability and mechanistic interpretability of the developed model. Thus, a huge number of investigations are currently directed towards the introduction of more appropriate validation approaches for more accurate and predictive QSAR model development. One important objective of QSAR modeling is to predict activity/property/toxicity of new chemical entities falling within the applicability domain of the developed models. The reliability of any QSAR model depends on the confident predictions of these new compounds based on applicability domain (AD) of the modeled compounds and therein arises the importance of the AD study (Ferreira, 2001).

In order to establish the scientific validity of a QSAR model and to facilitate its acceptance for regulatory purposes, the Organization for Economic Cooperation and Development (OECD) in its joint meeting (OECD, 2007) has agreed to five principles that should be followed during the construction of QSAR models. The OECD Principle 3 which defines the need of an AD expresses the fact that QSARs are unavoidably associated with restrictions in terms of the types of chemical structures, physicochemical properties and mechanisms of action for which the models can generate reliable predictions. The applicability domain of a QSAR model has been defined as the response and chemical structure space, characterized by the properties of the molecules in the training set. The developed QSAR model can predict a new compound truly only when it falls within the applicability domain of the developed model (Netzeva *et al*., 2005). Thus, to identify the interpolation (true prediction) or extrapolation (less reliable prediction) of query compounds is an important task for a QSAR model developer using the information of applicability domain (Netzeva *et al*., 2005).

Viewing the importance of AD in QSAR model validation, we herein focus to get an overview of different traditional as well as relatively new AD approaches used to judge the quality of the QSAR models. This book chapter will be helpful for the QSAR learners in order to have a clear idea on the principles of available AD approaches useful for judging the predictive quality of QSAR models.

BACKGROUND

A QSAR model is essentially valued in terms of its predictability, indicating how well it is able to predict the endpoint values of the compounds which are not used to develop the correlation. The models that have been suitably validated internally and externally can be considered reliable for both scientific and regulatory purposes (Golbraikh & Tropsha, 2002; Tong *et al*., 2004]. QSAR models should be validated according to the OECD principles for reliable prediction. A meeting of QSAR experts held in Setúbal, Portugal in March 2002 formulated guidelines for the validation of QSAR models, in particular for regulatory purposes (Jaworska *et al*., 2003). These principles were agreed by OECD member countries, QSAR and regulatory communities at the 37th Joint Meeting of the Chemicals Committee and Working Party on Chemicals, Pesticides and Biotechnology in November 2004. These principles are best possible summary of the most important points that are required to be addressed to find consistent, reliable and reproducible QSAR models (OECD, 2004). Five OECD principles to develop reliable models are:

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Principle 1: A defined endpoint;
Principle 2: An unambiguous algorithm;
Principle 3: A defined domain of applicability;
Principle 4: Appropriate measures of goodness-of fit, robustness and predictivity;
Principle 5: A mechanistic interpretation, if possible (OECD, 2004).

In order to provide guidance on the interpretation of the mentioned principles, the OECD has also provided a checklist (OECD, 2004). Thus, the current challenge in the process of development of a QSAR model is no longer in developing a model that is statistically sound to predict the activity within the training set, but in developing a model with the capability to predict accurately the activity of untested chemicals. In spite of high fitted accuracy and apparent mechanistic appeal, some published QSAR models fail rigorous validation tests, and, thus, may lack practical utility as reliable screening tools (Tropsha *et al*., 2003]. In summary, predictive power is one of the most important validation features of QSAR models. It can be defined as the ability of a model to predict accurately the target activity/ property/toxicity of compounds that are not used for model development.

In this context, QSAR model predictions are most reliable if they come from the model's applicability domain (AD) which is broadly defined under OECD principle 3. The OECD includes AD assessment as one of the QSAR acceptance criteria for regulatory purposes (OECD, 2004; OECD, 2007). The Setubal Workshop report (Jaworska et al., 2003) presented the following regulation for AD assessment: "The applicability domain of a (Q)SAR is the physico-chemical, structural, or biological space, knowledge or information on which the training set of the model has been developed, and for which it is applicable to make predictions for new compounds. The applicability domain of a (Q)SAR should be described in terms of the most relevant parameters, i.e., usually those that are descriptors of the model. Ideally the (Q)SAR should only be used to make predictions within that domain by interpolation not extrapolation. This depiction is useful for explaining the instinctive meaning of the "applicability domain" approach.

Two general principles to AD estimation have been proposed till date (Nikolova-Jeliazkova & Jaworska, 2005). The first one estimates the interpolation region based on the training set in the model's descriptor space. The second approach relies on similarity analysis according to the premise that a QSAR prediction is reliable if the compound is "similar" to the compounds in the training set. It is interesting to point out that "similarity" is a relative concept as different concepts of similarity are relevant to different endpoints. The similarity approach to AD assessment should be based on measurable key features relevant to the endpoint modelled precisely (Bender & Glen, 2004; Nikolova & Jaworska, 2004).

APPLICABILITY DOMAIN (AD)

The AD (Eriksson *et al*., 2003; Gramatica, 2007; Netzeva *et al*., 2005, Tetko *et al*., 2006) is a theoretical region in chemical space encompassing both the model descriptors and modeled response. In the development of a QSAR model, the domain of applicability of molecules plays a crucial role for estimating the uncertainty in the prediction of a particular compound based on how similar it is to the compounds used to build the model. Thus, the prediction of a modeled response using QSAR is valid only if the compound being predicted falls within the domain of applicability of the model since it is impossible to predict an entire universe of chemicals using a single QSAR model. Again, AD can be described as the physico-chemical,

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

structural or biological space, knowledge or information based on which the training set of the model is developed, and the model is applicable to make predictions for new compounds within the specific domain (Weaver & Paul Gleeson, 2008).

The selection method of the training and test sets has a significant effect on the applicability domain of constructed QSAR model. Thus, while splitting a dataset for external validation, the training set molecules should be selected in such a way that they span the entire chemical space for all the dataset molecules. In order to achieve successful predictions, a QSAR model should always be used for compounds within its applicability domain. Therefore, one can arguably agree that if a new compound falls outside of the AD of the training set molecules, its prediction is not reliable.

APPLICABILTY DOMAIN APPROACHES

There are various AD approaches that have been used by different groups of QSAR researchers worldwide. The most common approaches for estimating interpolation regions in a multivariate space include the following ones (Jaworska *et al*., 2005; Netzeva *et al*., 2005):

1. Ranges in the descriptor space.
2. Geometrical methods.
3. Distance-based methods.
4. Probability density distribution.
5. Range of the response variable.

The first four approaches are based on the methodology used for interpolation space characterization in the model descriptor space. On the contrary, the last one depends solely on response space of the training set molecules. A compound can be identified as out of the domain of applicability in a simple way, if:

1. At least one descriptor is out of range for the ranges approach, and
2. The distance between the chemical and the center of the training data set exceeds the threshold for distance approaches.

The threshold for all kinds of distance methods is the largest distance between the training set data points and the center of the training data set.

Along with the common approaches, Stanforth *et al*., (2007) proposed a cluster-based approach to evaluate the AD of any QSAR model. The method applies an intelligent version of the *k*-means clustering algorithm for modeling the training set as a compilation of clusters in the descriptors space. On the other hand, the test compounds of individual clusters are assigned a fuzzy membership from which an overall distance may be calculated to identify the AD of individual test compounds. A classification-based process involves calculation of regression residuals for identification of good and bad classes suggested by Guha & Jurs (2005).

All approaches are discussed thoroughly in this section enlightening their main theories to define the interpolation space as well as the threshold criteria for AD used in QSAR modeling. A summary of hypotheses of data distribution and interpolation space of most commonly used AD methods is discussed in Table 1.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

1. RANGES IN THE DESCRIPTOR SPACE APPROACH

1.1 Fixed or Probabilistic Boundaries/Bounding Box

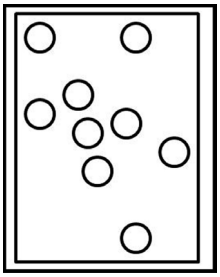
- **Theory:** The ranges of the individual descriptors are considered with a uniform distribution defining an n-dimensional hyper-rectangle developed on the basis of the maximum and minimum values of each descriptor used to build the model with sides parallel to the coordinate axes.
- **Criteria:** This method is simply based on boundary created by individual descriptor ranges of the training set compounds from which the QSAR model is developed. As mentioned earlier, this is based on the highest and lowest values of X variables (descriptors to develop the QSAR model) and Y variable (response for which the QSAR equation is created) of the training set. Any test set compounds, which are not present in any of these particular ranges, are considered out of the AD and their predictions are less reliable (Jaworska *et al*., 2005; Seber, 1984).
- **Drawbacks:**
  - For non-uniformly distributed data, the approach encloses considerable empty space,
  - Empty regions in the interpolation space cannot be recognized, as only the descriptor ranges are considered, and
  - Correlation between descriptors cannot be taken into account (Jaworska *et al*., 2005).

For the 'bounding box' method, the domain of applicability is the smallest axis-aligned rectangular box containing all the data points presented in Figure 1.

1.2 Principal Component Ranges/PCA Bounding Box

- **Theory:** Principal components convert the original data into a new orthogonal coordinate system by the rotation of axes and enable to correct for correlations among descriptors. Newly formed axes are defined as Principal Components (PCs) representing the maximum variance of the total dataset. The points between the lowest value and the highest value of each PC define an M-dimensional (*M* is the number of significant components) hyper-rectangle with sides parallel to the PCs (Nikolova-Jeliazkova & Jaworska, 2005).
- **Criteria:** This hyper-rectangle AD also includes empty spaces depending on uniformity of data distribution (Nikolova-Jeliazkova & Jaworska, 2005). Combining the bounding box method with principal component analysis (PCA) can overcome the setback of correlation between descriptors

Figure 1. Bounding box plot



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 1. Summary of hypothesis of most commonly used AD methods

Method		Formula/Equation	Hypothesis on Data Distribution	Shape of the AD Space
Range based		$d(x,y) = x - y$	Uniform	Boundary created by individual descriptor of the training set compounds
Geometric (Convex hull)		-	Uniform and smallest axis-aligned convex region containing all the data points	Convex
Distance-based	Mahalanobis distance/ Hotelling T^2/leverage	$D_M(x,\mu) = \left(x-\mu\right)^T \ \Sigma^{-1}\left(x-\mu\right)$	a) Normal, b) arbitrary variances, and c) arbitrary correlation. Mean μ, Covariance matrix Σ $p\left[x\right] = \frac{1}{\left(2\pi\right)^{N/2}\left\|\sum\right\|^{\frac{1}{2}}} \exp\left\langle -\frac{1}{2} D_M\left(x,\mu\right)\right\rangle$	Ellipses (hyper-ellipses)
	Euclidean distance	$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$ *n*=total number of observations for which a pair wise comparison is performed (x_i and y_i), *i* is the number of descriptors	a) Normal, b) equal variances, and c) uncorrelated variables Mean μ, unit covariance matrix $p\left[x\right] = \frac{1}{\left(2\pi\right)^{N/2}} \exp\left\langle -\frac{1}{2} D_E^2\left(x,\mu\right)\right\rangle$	Spheres (hyper-spheres)
	City block	$d(x,y) = \sum_{i=1}^{n}\left\|x_i - y_i\right\|$	Multivariate uniform	Rectangle
Probability density distribution		$\varphi = \frac{1}{\sqrt{\left(2\pi\right)}s} . \exp\left[\frac{-1}{\left(2s^2\right)\left(x_i - x_j\right)^2}\right]$ where, Φ (x_i and x_j) is the potential induced on x_j by x_i and width of the curve is defined by smoothing parameter *s*. The cut off value associated with Gaussian potential functions, namely f_p, can be calculated by: $f_p = f_i + \left(q-j\right)\left(f_{j+1} - f_{\ j}\right)$ where, $q = p \times {}^{n}\!/_{100}$, p is the percentile value of probability density, *n* is the number of compounds in the training set and *j* is the nearest integer value of q.	1. Parametric methods use standard distributions such as Gaussian and Poisson distributions. 2. Non-parametric techniques permit the estimation of probability density solely from data.	Captures actual data distribution and no reference data point are considered. For each data point, the method calculates the probability of belonging to the set.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

but the issue regarding empty regions within the interpolation space still remains. It is interesting to point out that the empty space is smaller than the hyper-rectangle in the original descriptor ranges. The most important criterion of this approach remains in the selection of appropriate number of components (Jaworska *et al*., 2005, Worth *et al*., 2005). The PCA bounding box has been used to analyze the AD of the KOWWIN model.

The AD in principal component space is presented in Figure 2. The QSAR model training set is represented by the biggest circle. The predictions of query molecules within the training space are considered reliable. Query compounds which are located outside the training model space would be expected to be less reliably predicted.

1.3 TOPKAT Optimal Prediction Space

- **Theory:** Variation of PCA is implemented by the Optimum Prediction Space (OPS) from TOPKAT OPS (2000). In the PCA approach, instead of the standardized mean value, the data are centered around the mean of individual parameter range ($[x_{max}–x_{min}]/2$). Thus, it establishes a new orthogonal coordinate system which is known as OPS coordinate system. Here, the basic process is same, extracting eigenvalues and eigenvectors from the covariance matrix of the transformed data. The OPS boundary is defined by the minimum and maximum values of the generated data points on each axis of the OPS coordinate system.
- **Criteria:** The Property Sensitive object Similarity (PSS) is implemented in the TOPKAT as a heuristic solution to replicate the data set's dense and spare regions, and includes the response variable (y). Accuracy of prediction between the training set and queried points is assessed by the PSS. Similarity search method is used to evaluate the performance of TOPKAT in predicting the effects of a chemical that is structurally similar to the training structure.

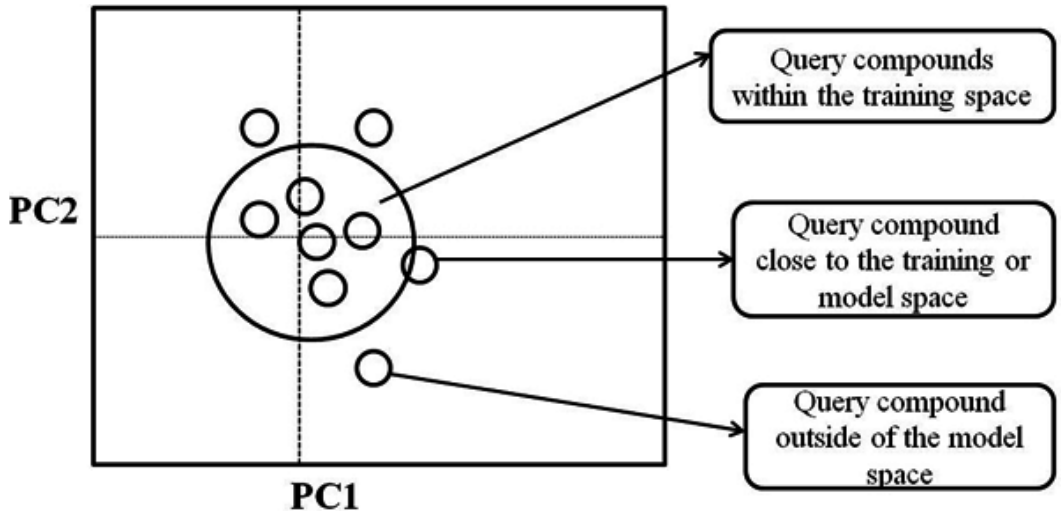
2. GEOMETRICAL METHODS

2.1 Convex Hull

- **Theory:** This approach estimates the direct coverage of an n-dimensional set using the convex hull calculation (Preparata & Shamos, 1991). Convex hull calculation is a computational geometry problem which is performed based on complex but efficient algorithms. The approach recognizes the boundary of the dataset considering the degree of data distribution.
- **Criterion:** Interpolation space is defined by the smallest convex area containing the entire training set.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 2. A graphical illustration of the AD in principal component (PC) space



- **Drawbacks:**
  - Implementing a Convex Hull can be challenging with increasing data complexity, this means that an increase in dimensions contributes to the order of complexity. Convex hull calculation is efficient for two and three dimensions. The complexity swiftly amplifies in higher dimensions. For n points and d dimensions, the complexity is of order O which can be defined as $O = [n^{[d/2]+1}]$ (Preparata & Shamos, 1991),
  - The approach only analyses the set boundaries without considering the actual data distribution, and
  - This approach cannot identify the potential internal empty regions within the interpolation space (Jaworska *et al*., 2005).

For the 'convex hull' method, the domain of applicability is the smallest axis-aligned convex region containing all the data points presented in Figure 3.

3. DISTANCE-BASED METHODS

Distance-based methods are mainly founded on the 'distance-to-centroid' principle. The ellipsoidal region is centered on the dataset grand mean and has its principal axes based on the eigenvectors of the dataset variance/covariance matrix. In Figure 4, a graphical illustration of 'distance-to-centroid' is presented.

There are various types of distance-based methods which can be used for the determination of AD of QSAR models. We have tried to represent the most commonly used distance-based AD approaches in the following sections.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 3. Convex hull plot

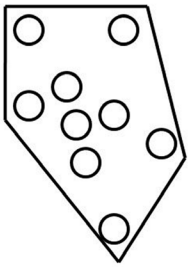
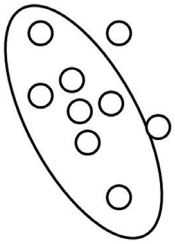


Figure 4. Distance-to-centroid plot



3.1 Leverage Approach

- **Theory:** The leverage (*h*) of a compound in the original variable space is calculated based on the HAT matrix as:

$$H = (X^T(X^TX)^{-1}X). \quad (1)$$

In Equation (1), H is an [n x n] matrix that orthogonally projects vectors into the space spanned by the columns of X (Eriksson *et al*., 2003; Gramatica, 2007). The AD of the model is defined as a squared area within the ±3 band for standardized residuals (σ) and the leverage threshold is defined as $h^*=3(p+1)/n$, where p is the number of variables and n is the number of compounds. The leverage values (h) are calculated for each compound and plotted vs. cross-validated standardized residuals (σ) (Y-axis) referred to as the Willam's plot.

The leverage approach assumes normal data distribution. It is interesting to point out that when high leverage points fit the model well (having small residuals), they are called good high leverage points or good influence points. Those points stabilize the model and make it more accurate. On the contrary, high leverage points, which do not fit the model (having large residuals), are called bad high leverage points or bad influence points.

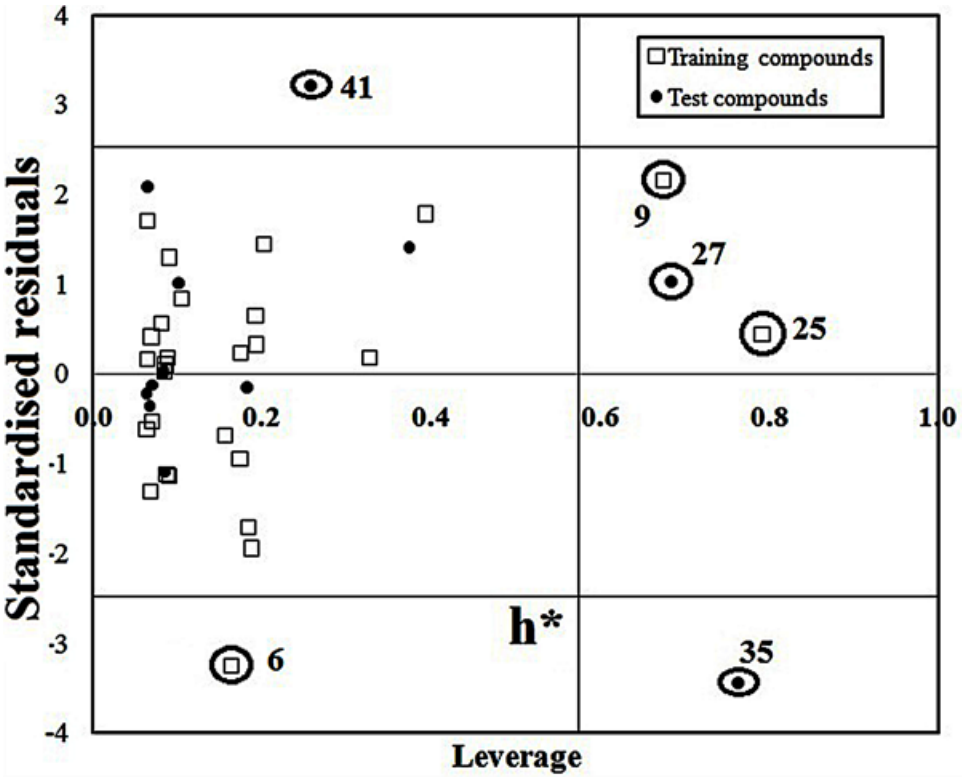
- **Criteria:** Graphically, the Williams plot (mostly used for the leverage approach) verifies the presence of response outliers and training set chemicals that are structurally very influential in determining model parameters. The data predicted for high leverage chemicals in the prediction set are extrapolated and could be less reliable. Here, we have tried to simplify the concept with a pictorial representation of Williams plot in Figure 5.

It is interesting to point out that this figure has been drawn based on arbitrary data to show the every possibility of AD assessment in training and test set compounds. Assessing Figure 5, one can draw the following conclusion:

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

1. Compound number 6 of the training set possesses standardized cross-validated residual value greater than -3σ unit, but it lies within the critical HAT (h*) value of 0.58. This compound may have flawed observed value and it can be identified as an outlier due to a wrong value of response variable. Therefore, compound number 6 is a response outlier.
2. Compounds with serial numbers 9 and 25 belonging to the training set are characterized by high leverage values, but they lie within the fixed ±3σ limit of the ordinate. The mentioned chemicals are not response outliers (not Y outliers), but influence the applicability domain to span over larger area and hence can be marked as influential chemicals or X outliers.
3. Compound number 35 belonging to the test set is wrongly predicted having standardized cross-validated residual value greater than -3σ unit) as well as completely outside of the AD as defined by HAT vertical line (higher leverage value than the critical value).
4. On the contrary, one of the test set compounds (41) belongs to AD as it lies within the critical HAT value but the prediction for this compound is not precise (greater than 3σ unit).
5. The test set compound with serial number 27 is within the ±3σ limit of the ordinate, but possess higher leverage value (h > h*). As we can see from Figure 5, this compound is surrounded by other two influential training set chemicals (serial numbers 9 and 25), therefore its prediction can be assumed to be somewhat reliable or dependable though it lies outside the AD.

Figure 5. Williams plot



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

3.2 Euclidean Distance

- **Theory:** This is one of the major distance based AD assessment approaches. Like other distance-based approaches, it calculates the distance from every other point to a particular point in the data set. The distance scores are calculated by the Euclidean distance norm. A distance score, d_{ij}, for two different compounds X_i and X_j can be measured by the Euclidean distance norm. The Euclidean distance can be expressed by the following equation:

$$d_{ij} = \sqrt{\sum_{k=1}^{m} (x_{ik} - x_{jk})^2} \,. \tag{2}$$

The mean distances of one sample to the residual ones are calculated as follows:

$$\overline{d_i} = \sum_{j=1}^{n} d_{ij} \Big/ n-1, \tag{3}$$

where, i=1,2,….,n.

The mean distances are then normalized within the interval of zero to one. It is applicable only for statistically independent descriptors (Jaworska *et al*., 2004).

- **Criteria:** Compounds with distance values adequately higher than those of the most active probes are considered to be outside the domain of applicability. The mean normalized distances are measured for both training and test set compounds. The boundary region created by normalized mean distance scores of the training set are considered as the zone of applicability domain for test set compounds. If the test set compounds are inside the domain/area covered by the training set compounds, it means that these compounds are inside the applicability domain, otherwise not.

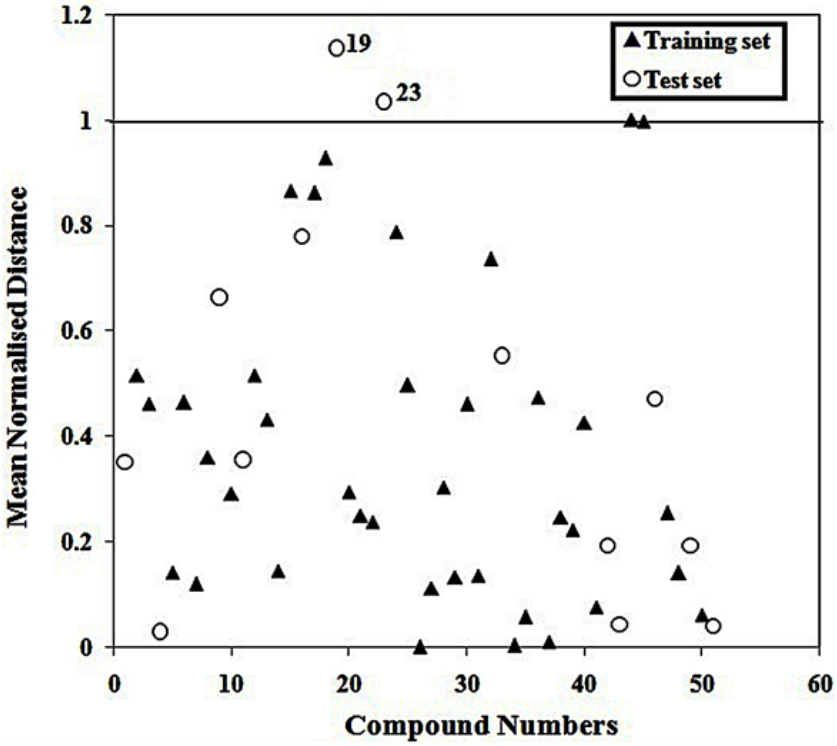
An example of Euclidean plot is presented in Figure 6. Based on the assessment of the Euclidean plot, one can conclude that test set compounds 19 and 23 are outside of the applicability domain created by the training set compound from which the QSAR model has been developed.

3.3 Mahalanobis Distance

- **Theory:** The Mahalanobis distance approach considers the distance of an observation from the mean values of the independent variables but not taking the impact on the predicted value. It offers one of the unique and simple approaches for identification of outliers. Mahalanobis distance is unique because it automatically takes into account the correlation between descriptor axes (Hair *et al*., 2005).
- **Criterion:** The threshold value does not have any rule of thumb. However, observations with values much higher than those of the remaining ones may be considered to be outside the domain.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 6. Euclidean distance plot for AD study of an arbitrary QSAR model



3.4 City Block Distance

City-block distance is simply the summed difference across dimensions and is calculated from the following equation:

$$d(x,y)=\sum_{i=1}^{n}\left|x_i-y_i\right|. \tag{4}$$

It examines the absolute differences between coordinates of a pair of objects (x_i and y_i). City-block distance assumes a triangular distribution. The method is particularly useful for the discrete type of descriptors. It is used only for training sets which are uniformly distributed with respect to count-based descriptors (or fragment mapping counts) (Jaworska *et al*., 2004).

3.5 Hotelling T^2 Test

- **Theory:** The Hotelling T^2 method is a multivariate student's *t* test and proportional to leverage and Mahalanobis distance approach. It assumes a normal data distribution like the leverage approach (Hair *et al*., 2005). The method is used to evaluate the statistical impact of the difference on the

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

means of two or more variables between two groups. Hotelling T^2 corrects for collinear descriptors through the use of the covariance matrix. Hotelling T^2 measures the distance of an observation from the center of a set of X observations. A tolerance volume is derived for Hotelling T^2.

- **Criterion:** Based on the *t* value, the significant compounds within the domain are determined.

3.6 K-Nearest Neighbors Approach

- **Theory:** The approach is based on similarity search for a new chemical entity with respect to the space created by the training set compounds. The similarity is identified by finding the distance of a query chemical from nearest training compound or its distances from *k*-nearest neighbors in the training set. Thus, similarity to the training set molecules is significant for this approach in order to associate a query chemical with reliable prediction (Sheridan *et al*., 2004). Like other methods, descriptors are used for calculate the similarity between training and test molecules.
- **Criterion:** If the calculated distance values of test set compounds or query molecules are within the user defined threshold set by the training set molecules, then the prediction of these compounds are considered to be reliable.

How One Can Find the Similarity to the Molecules of the Training Set

1. Considering all the descriptors, the mean similarity of a test set molecule to the most similar *k* members of the training set can be found. A variation of this is to calculate this similarity using only the subset of descriptors in the training set.
2. The similarity of the test set molecule to the descriptor centroid of the training set.

How to Decide the Neighboring Test or Query Compounds in the Training Set

1. The simplest method is to pick a similarity cutoff above which the test set molecule and a training set molecule can be considered "neighbors" and count the molecules that exceed that cutoff.
2. Another complicated approach is explained by Sheridan et al. (2004) which demonstrated that "A more complicated approach is to define a falloff function f(Similarity) that defines how much of a neighbor a training set molecule is as a function of similarity to the test set molecule. The number of neighbors is the sum of f (Similarity) over all members of the training set. For instance, we use a linear function where a molecule is 1.0 of a neighbor at similarity 1.0 and 0.0 of a neighbor at a similarity 0.6; anything between is linearly interpolated."

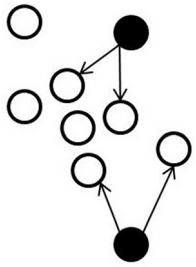
A k-nearest neighbors plot is presented in Figure 7, where the distance of a point to the domain is taken as the average distance to the k- nearest data points.

3.7 DModX (Distance to the Model in X-Space)

- **Theory:** This is another important distance based AD finding approach for QSAR models. This approach was developed by Wold *et al*., (2001) and usually applied for partial least squares (PLS) models. The basic theory lies in the residuals of Y and X which are of diagnostic value for the quality

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 7. K-nearest neighbors plot



of the model. As there are number of X-residuals, one needs a summary for each observation. This is accomplished by the residual standard deviation (SD) of the X-residuals of the corresponding row of the residual matrix E. As this SD is proportional to the distance between the data point and the model plane in X-space, it is also usually called DModX (distance to the model in X-space). Here, X is the matrix of predictor variables, of size (N*K), Y is the matrix of response variables, of size (N*M) and E is the (N*K) matrix of X-residuals, N is number of objects (cases, observations), k is the index of X-variables (k = 1, 2, . . ., K) and m is the index of Y-variables (m = 1, 2, . . ., M).

- **Criteria:** A DModX value larger than around 2.5 times the overall SD of the X residuals (corresponding to an F-value of 6.25) indicates that the observation is outside the applicability domain of the model. In the DModX plot, the threshold line is attributed as D-critical line and this plot can be easily drawn in SIMCA-P software for any PLS model (SIMCA-P 10.0, 2002).

An example of DModX plot is presented in Figure 8 where the plot is constructed at 95% confidence level by SIMCA-P. The DModX values of 26 test compounds are within the critical value of 1.986. Chemical **10** along with the compound numbers **3**, **25**, **53** and **73** are outside the evaluated critical value indicating that these compounds are outside of the applicability domain of the developed model.

3.8 Tanimoto Similarity

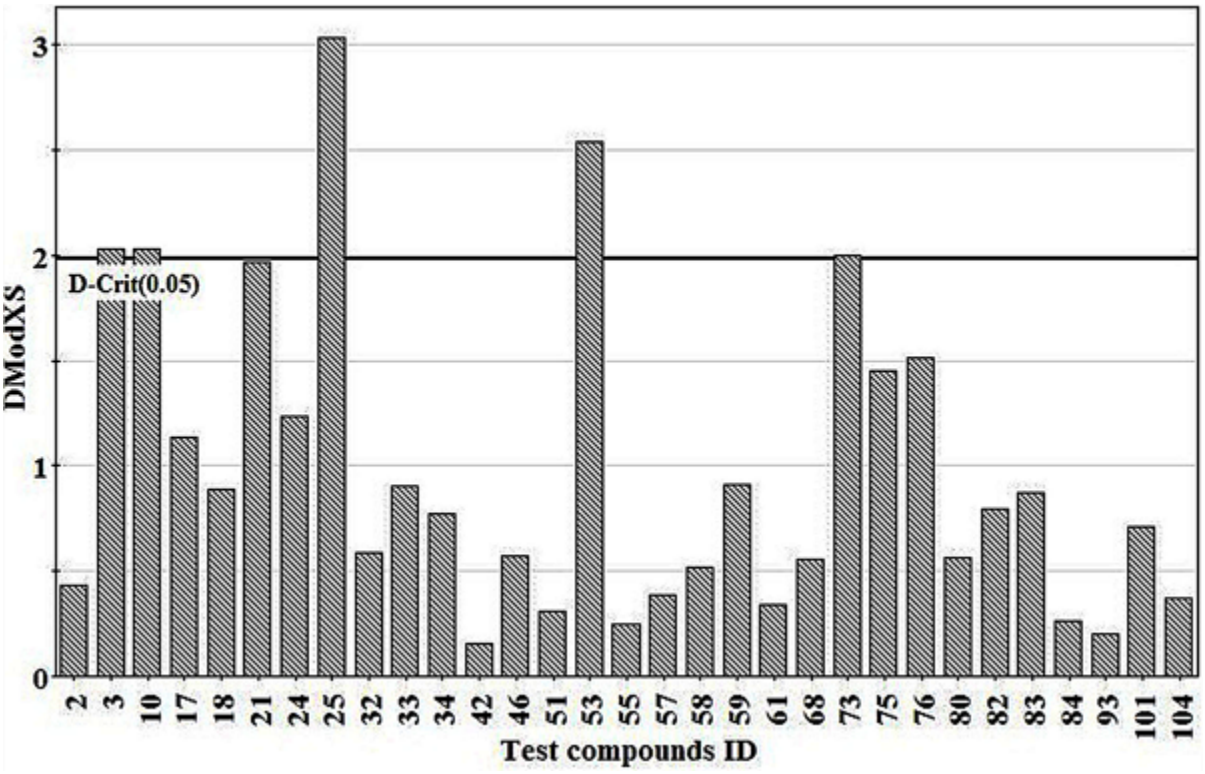
The Tanimoto index is a measure of similarity between two molecules based on the number of common molecular fragments (Tetko *et al*., 2008). In order to calculate the Tanimoto similarity, all unique fragments of a particular length in two compounds are calculated. The Tanimoto similarity between the compounds *J* and *I* is defined as:

$$TANIMOTO\left(J,K\right)=\frac{\sum_{i=1}^{N}\left(x_{J,i}\cdot x_{K,i}\right)}{\sum_{i=1}^{N}\left(x_{J,i}\cdot x_{J,i}\right)+\sum_{i=1}^{N}\left(x_{K,i}.x_{K,i}\right)-\sum_{i=1}^{N}\left(x_{J,i}.x_{K,i}\right)} \quad (5)$$

where, *N* is the number of unique fragments in both the compounds, $x_{J,i}$ and $x_{K,i}$ are the counts of the *i*-th fragment in the compounds *J* and *K*. Based on equation 3, the distance between two compounds *J* and *K* is "*1– TANIMOTO(J, K)*" and the distance of a compound to the model is the minimum distance between the investigated compound and compounds from the training set of the model.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 8. DModX plot for PLS based QSAR model



3.9 Standard Deviation of the Ensemble Predictions (STD)

In order to estimate the uncertainty of the model, standard deviation of the predictions obtained from an ensemble of models can be utilized. The basic principle behind this idea is very simple and it depicts that if different models give significantly different predictions for a particular molecule, then the prediction for this compound is more likely to be unreliable. The sample standard deviation can be ideally used as an estimator of model uncertainty.

As an example, consider that $Y(J) = \{y_i(J), i=1..N\}$ is a set of predictions for a compound J given by a set of N trained models, the corresponding distance to model STD can be defined by the following equation:

$$d_{STD}(J) = stdev\left(Y\left(J\right)\right) = \sqrt{\frac{\sum\left(y_i\left(J\right) - \bar{y}\right)^2}{N-1}} \,. \tag{6}$$

Based on the type of an ensemble used to evaluate the standard deviation, the STD approach can be classified into several subtypes. (a) Consensus STD or CONS-STD for models developed based on different machine learning techniques, (b) Associative neural networks or ASNN-STD for an ensemble of neural network models, and (c) BAGGING-STD for an ensemble of models formed utilizing the bagging technique. The distance based method STD has been verified to provide outstanding results for discrimination of highly accurate predictions for regression models (Manallack *et al*., 2003; Tetko *et al*., 2008).

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

A graphical illustration of the STD approach has been given in Figure 9. Here, reliable prediction has a low prediction spread and unreliable prediction has higher prediction spread.

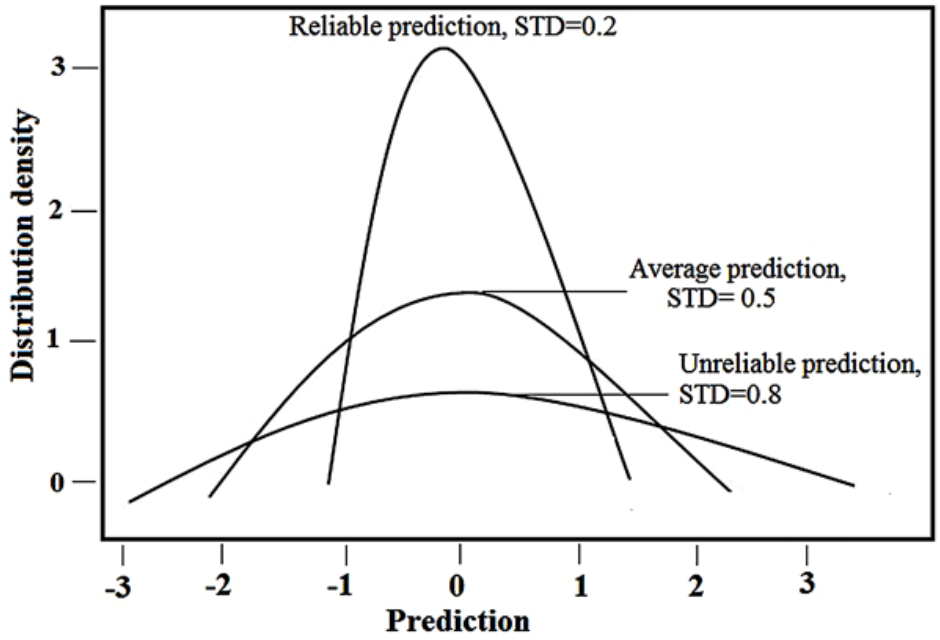
3.10 Correlation of Prediction Vectors (CORREL)

The basic theory of this approach is based on the correlation of vectors of ensemble's predictions for the target compound and compounds from the training set (Tetko, 2008). Similar to the STD method, this measure is relevant only for ensembles of models. More specifically, CORREL measure for the target compound *J* is calculated according to the following expression:

$$d_{CORREL}\left(J\right) = 1 - \max_{i=1\ldots n}\left[corr\left(y\left(\vec{T}_i\right), y\left(\vec{J}\right)\right)\right]. \tag{7}$$

In Equation (7), $y\left(\vec{T}_i\right)$ and $y\left(\vec{J}\right)$ define the vectors of ensemble's predictions for the training set compound *Ti* and the target compound *J*, *corr* is Spearman rank correlation coefficient between the two vectors and *N* is the number of compounds in the training set. A low value of CORREL (that indicates the high Spearman correlation coefficient) specifies that for target compound *J,* there is a compound *T* from training set for which predictions of the ensemble of models are strongly correlated. If a compound *T* has the same descriptors as the target compound *J*, then predictions of models will be identical for both molecules and thus resulted *CORREL(J)* will be 0. Compounds having high correlation coefficient values are considered to be "closer to the model". The Spearman correlation outperformed several al-

Figure 9. STD graph for AD estimation



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

ternative measures (as e.g. Euclidean distance and Pearson correlation coefficient) due to its higher accuracy and practical requirements to store ensemble predictions for all training set molecules (Tetko & Tanchuk, 2002).

3.11 Concordance of a Classification Ensemble

The highest number or the relevant percentage of the models that produce the same prediction can be used as a measure of the concordance of an ensemble (Chen *et al*., 2005). As an example, if there are 5 models that give predictions {a; b; a; a; a}, then the concordance is 4 (or 80%). The measure that is opposite to the concordance (*i.e.* a- CONCORDANCE) can be used as a distance to model. The idea behind this method is similar to that of the standard deviation (STD) distance to model approach, but is adapted for classification models and qualitative predictions.

4. PROBABILITY DENSITY DISTRIBUTION

- **Theory:** Probability density distribution is one of the most advanced approaches for defining AD, based on estimating the probability density function for the given data. The approach is classified into two classes. The first one is a parametric method which assumes a standard distribution such as Gaussian and Poisson distributions, and the second one is a non-parametric method which does not rely on such assumptions considering the data distribution (Jaworska *et al*., 2005; Netzeva *et al*., 2005).

These methods are executed by estimating probability density of the compounds followed by identifying Highest Density Region which consists of a known fraction (as set as user input) from the total probability mass (Netzeva *et al*., 2005). A potential is formed for each compound in the training set such that it is highest for that molecule and decreases with increasing distance. Once the potential is evaluated for all the compounds, global potential is obtained by summing up the individual potentials thus indicating the probability density (Forina *et al*., 1991; Jouan-Rimbaud *et al*., 1999).

The most important feature of these methods is their capability to identify the internal empty space. The actual data distribution can be revealed by generating concave regions around the interpolation space borders.

- **Criteria:** Here, Gaussian function is demonstrated. Given two molecules x_i and x_j, it can be determined as below:

$$\varphi = \frac{1}{\sqrt{(2\pi)}s}.\exp\left[\frac{-1}{\left(2s^2\right)\left(x_i - x_j\right)^2}\right] \tag{8}$$

where, Φ (x_i and x_j) is the potential induced on x_j by x_i and width of the curve is defined by smoothing parameter s. The cut off value associated with Gaussian potential functions, namely f_p can be calculated by methods based on sample percentile (Jouan-Rimbaud *et al*., 1999):

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

$$f_p = f_i + \left(q - j\right)\left(f_{j+1} - f_j\right) \tag{9}$$

with $q = p \times {}^{n}\!/_{100}$, where p is the percentile value of probability density, n is the number of compounds in the training set and j is the nearest integer value of q. Test compounds with potential function values lower than the set threshold are considered outside the AD.

5. RANGE OF THE RESPONSE VARIABLE

Based on the distribution plot or considering the range of the response variable, compounds can be identified as outside the AD. In a series of compounds, if the response value of a particular test compound is largely different from the mean response value of the training set compounds, then the compound can be considered as out of the AD.

6. MISCELLANEOUS APPROACHES

Along with the above-mentioned strategies to determine the AD of constructed QSAR models, there are several miscellaneous approaches introduced in the recent times. Here, we have tried to discuss these approaches briefly. The use of these approaches has been very limited compared to the classical AD methods. The below-mentioned methods are especially used for interpolation space characterization in the model descriptor space.

6.1 Stepwise Based Approach

- **Theory:** Stepwise based approach for evaluating the AD of a QSAR model is proposed by Dimitrov *et al*. (2005). This approach is useful for reflecting mechanistic rationality and transparency of the QSAR model. This approach is processed in four stages in a chronological manner. The first stage "General requirements domain" consists of checking of the variation of the physicochemical properties of the query chemicals in the training set compounds. The second stage "Structural domain" defines the structural similarity which is found within the chemicals that are correctly predicted by the model. The third stage "Mechanistic domain" takes into account of the mechanistic interpretation of the modeled phenomenon to define the domain of applicability. In this third step, the model domain merges the reliability of specific functional groups assumed to cause the effect and the domain of explanatory variables. Here, the latter is defined by the parameter interpolation space. In the fourth and final stage "Domain of metabolic stimulator", if stimulated metabolism of chemicals is a part of the developed QSAR model then the dependability of simulated metabolism is considered in assessing the reliability of the predictions.
- **Criterion:** An external compound is compulsory to satisfy all the conditions specified within the mentioned four stages to be considered within the AD. The approach is considered as a rigorous one as a chemical is assessed for similarity, metabolic and mechanistic check to address the reliability of predictions and permitting a better assessment of model's AD (Dimitrov *et al*., 2005).

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

- **Drawback:**
  - In order to determine the AD of the QSAR model, one would need a lot of information regarding the structural attributes of compounds, mechanism of action for particular response as well as their metabolic activation. In regular practice of developing QSAR models, this is lacking in major cases.
  - The method is time consuming as well as critical as a lot of variables are involved in this AD assessment approach.

6.2 Decision Trees and Decision Forests

This approach identifies the AD in terms of prediction confidence and domain extrapolation based on the consensus prediction of Decision Trees (DT) and Decision Forests (DF). The basic principle is to minimize the overfitting which can be attained by merging the DTs and keeping the differences within different DTs to maximum possible. Predictions from all the combined DTs are averaged in order to find the prediction confidence for a particular molecule, while domain extrapolation provides the prediction precision for that compound outside the training space (Netzeva et al., 2005; Tong et al., 2003, 2004).

6.3 Kernel-Based Applicability Domain

Machine learning tools are becoming an important method to evaluate the molecular activities/properties without the help of *in vitro* experiments. The major drawback of these methods is that most of them do not provide any information using the knowledge contained in the model though they are predicted adequately by the model. As a result, in machine learning based QSAR modeling, the estimation of the reliability of a model-based prediction is an important question (Fechner *et al*., 2009).

Most machine learning approaches for QSAR heavily depend on a vectorial representation of the molecules. Thus, the applicability domain is expressed as a subspace of the vector space with one dimension for individual descriptor used. However, this vectorial concept cannot be directly applied to kernel-based techniques like support vector machines. So, these methods have to rely on an implicit feature space which is only defined by the applied kernel similarity and with unknown dimensions. Therefore the domain of applicability of a kernel-based model has to be defined by means of the kernel. This also permits us to employ the structured similarity measures like the Optimal Assignment Kernel and its extension, instead of a numerical encoding. This method makes it possible to depict the complex chemical structure of many drugs better than the methods which are processed by using descriptors.

The concept of kernel density estimation has incorporated additional information in a trained model. The added information can be achieved by using a weighted average kernel similarity of a predicted molecule to the training data set. The weights can be attained either by utilizing the knowledge contained in the learned model or by methods that describe the feature space structure using the kernel (Fechner *et al*., 2009).

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

6.4 Intelligent *K*-Means Clustering

An intelligent version of the k-means clustering algorithm-based measure of distance-to-domain for QSAR models was designed by Stanforth *et al*., (2007) to obviate the difficulties of the existing methods (like convex hull, bounding box etc.) in terms of primitiveness and computational complexity. This measure combines the modeling of the training set as a collection of the intelligent k-means clusters algorithm in the descriptor space with a new interpretation of a conventional optimization criterion in fuzzy clustering which leads to a modified harmonic mean measure. A test compound is assigned fuzzy membership of each individual cluster, from which an overall distance may be calculated. Stanforth *et al*., (2007) demonstrated that this approach is more stable than existing methods and more indicative of the prediction error.

k-Means clustering is prepared based on the approximation of each point in a dataset by the centroid of that point's cluster. This estimation can be quantified by decomposing the data scatter into contributions that are explained and unexplained by the cluster model as presented in the following equation (Mirkin, 2005):

$$\sum_{i\in T} x_i \cdot x_i = \sum_{k=1}^{k} N_k c_k \cdot c_k + \sum_{k=1}^{k} \sum_{i\in C_k} \left(x_i - c_k\right)\cdot\left(x_i - c_k\right) . \tag{10}$$

In Equation (10), $\{x_i: i \in T\}$ is the training set and the k clusters C_k have sizes N_k and centroids c_k, respectively. This suggests an optimization of the so-called square-error k-means criterion

$$\sum_{k=1}^{k} \sum_{i\in C_k} \left(x_i - c_k\right)\cdot\left(x_i - c_k\right) . \tag{11}$$

Equation 11 holds for any inner product in the descriptor space.

k-Means clustering requires an initialization step to indicate the number of clusters k and the preliminary positions of their centroids. The initialization of the intelligent k-means algorithm has been achieved by the principle of Anomalous Pattern Clustering (APC) (Mirkin, 2005; Smellie, 2004). The principle can be demonstrated in the following way.

1. Two initial centroids are specified as follows: the first one is considered as the data gravity centre where components are grand means of the corresponding components in the entire dataset, and the second is an entity which is the furthermost away from the gravity centre.
2. It is interesting to point out that the first point (grand mean) centroid never changes, so that the only changeable centroid is the second one.

The computational complexity proved to be quite acceptable for this method. Assessment of the domain of applicability using clusters affords a much faster and easier calculation than k-NN. Not only that, this one is more prudent in terms of additional stored model parameters, without negotiating the acuity of the measure.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

PRACTICAL EXAMPLES OF DETERMINATION OF AD

In order to show an example of determination of the AD for a regression based QSAR model, we have shown in Table 2 a set of hypothetical X-Y data and a correspoding QSAR model developed with multiple linear regression (MLR) tool. The model consists of three descriptors. Here, we have shown the calculation and determination of parameters most commonly used in simple AD approaches like fixed or probabilistic boundaries (bounding box), range of the response variable, leverage method and Euclidean distance based method. All these methods can be used by any QSAR researchers without help of any professional software.

Table 2. A simple example demonstrating calculation of applicability domain for a regression-based QSAR model using different methods such as fixed or probabilistic boundaries (bounding box), leverage method, and Euclidean distance based approaches

Compound ID	Y* (Observed)	X_1*	X_2*	X_3*	Y (Calculated)	h[a]	Residual	Standardized Residual	Distance Score[b]	Mean Distance[c]	Normalized Mean Distance
1	3.45	1.65	1	1.65	3.46	0.41	-0.01	0.01	9.13	0.83	0.68
3	3.14	1.63	1.5	1.09	2.73	0.17	0.41	2.01	5.54	0.50	0.02
5	2.82	2.02	1.5	1.35	2.67	0.26	0.15	0.76	6.37	0.58	0.17
7	2.64	2.05	1.5	1.37	2.67	0.29	-0.03	-0.08	6.59	0.60	0.21
9	2.29	1.83	1.5	1.22	2.70	0.17	-0.41	-1.94	5.56	0.51	0.02
10	2.2	1.9	2	0.95	1.97	0.38	0.23	1.15	7.70	0.70	0.41
11	2.15	1.33	2	0.67	2.06	0.50	0.09	0.46	9.12	0.83	0.68
13	1.74	1.54	2	0.77	2.03	0.34	-0.29	-1.36	7.99	0.73	0.47
15	3.45	1.91	1	1.91	3.42	0.66	0.03	0.19	10.87	0.99	1.00
16	2.51	1.66	1.5	1.11	2.73	0.16	-0.22	-1.01	5.47	0.50	0.00
17	2.87	1.1	1.5	0.73	2.82	0.66	0.05	0.31	9.34	0.85	0.72
Test Set											
Compound ID	**Y* (Observed)**	**X_1***	**X_2***	**X_3***	**Y (Predicted)**	**h**	**Residual**	**Standardized Residual**	**Distance Score**	**Mean Distance**	**Normalized Mean Distance**
2	3.2	1.9	1	1.9	3.42	0.39	-0.22	-1.02	10.78	0.98	0.99
4	2.87	1.22	1.5	0.81	2.80	0.33	0.07	0.40	8.32	0.76	0.53
6	2.81	1.78	1.5	1.18	2.71	0.14	0.10	0.53	5.51	0.50	0.01
8	2.49	1.61	1.5	1.07	2.74	0.16	-0.25	-1.14	5.67	0.52	0.04
12	2.01	1.96	2	0.98	1.96	0.30	0.05	0.28	7.92	0.72	0.45
14	3.51	1.88	1	1.88	3.42	0.38	0.09	0.46	10.62	0.97	0.96

*Y is the response variable; X_1, X_2 and X_3 are the descriptors involved in the QSAR model, h is defined as the leverage value, [a]calculated from equation 1, [b]calculated from equation 2 and [c]calculated from equation 3.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

1. **Leverage Approach:** Based on the calculated leverage (h) and standardized residual values, we have constructed Williams plot (Figure 10). Here, the standardized residual (SR) value is calculated based on the following equation:

$$SR = \frac{\mathrm{Re}\,sidual\ \ value\ -\ Average\ \mathrm{Re}\,sidual\,value_{(Training)}}{S\tan dard\ \ deviation\,of\ \mathrm{Re}\,sidual\,value_{(Training)}}. \quad (12)$$

Then Williams plot is constructed taking standardized values in Y-axis and leverage values in X-axis. As the number of descriptors is three and number of compounds in the training set is eleven, the critical leverage (h*) value is = [3*(3+1)]/11=1.09. Taking into consideration the criteria described in section 3.1, one can firmly conclude that based on the leverage approach for the studied QSAR model, all the query compounds are within the AD.

2. **Euclidean Distance-Based Approach:** First, Euclidean distance scores are calculated (based on Equation 2) for the total set of compounds and then the mean distance scores (based on Equation 3) are calculated. The mean distances are then normalized within the interval of zero to one. From the normalized mean distances, it is clear that not a single test compound resides outside of the AD. The Euclidean plot (Figure 11) is then developed taking the mean normalized distance in Y-axis and compound numbers in X-axis.
3. **Fixed or Probabilistic Boundaries (Bounding Box) and Range of the Response Variable:** Both approaches are strictly dependent on the boundary created by independent and response variables of the training set compounds. Scrutinizing the descriptor values for the training and test sets, one can easily identify that not a single test set compound falls outside the AD boundary created by the independent variables. On the contrary, only one compound (14) can be considered outside of the AD boundary when someone is taking in consideration the response variable range. Therefore, one can conclude that considering both methods, only one compound (14) resides outside the AD though it is lying very close to the boundary created by response variable.

While taking into consideration the above mentioned methods for the identification AD for a single QSAR model, one may remember that no single method of AD determination can be considered as the universally the best one. It may also be noted that the results of one specific method may not match with those from another approach as each method is based on different statistical algorithm for the determination of AD. Thus, multiple methods should be applied before coming to a final conclusion regarding AD of a QSAR model.

FUTURE RESEARCH DIRECTION

It is generally believed that no single AD approach can be relied as universally ideal one to identify the interpolation region for any QSAR model. Extensive and fruitful research is being carried out on the development of some new approaches or modification of already existing approaches to make them more acceptable and reliable ones. The following points should be considered in future research regarding AD studies.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 10. Williams plot for the arbitrarily developed QSAR model

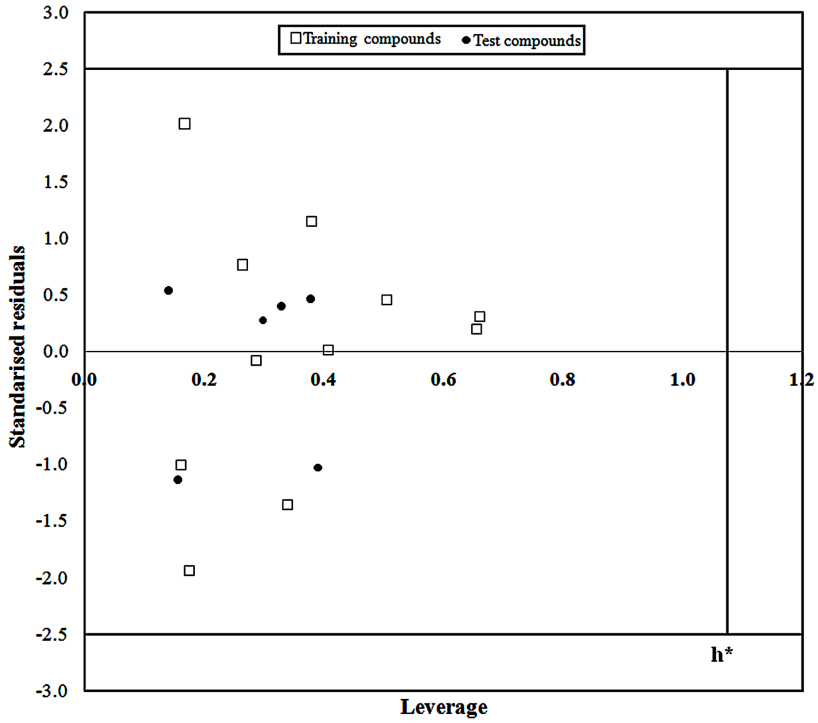
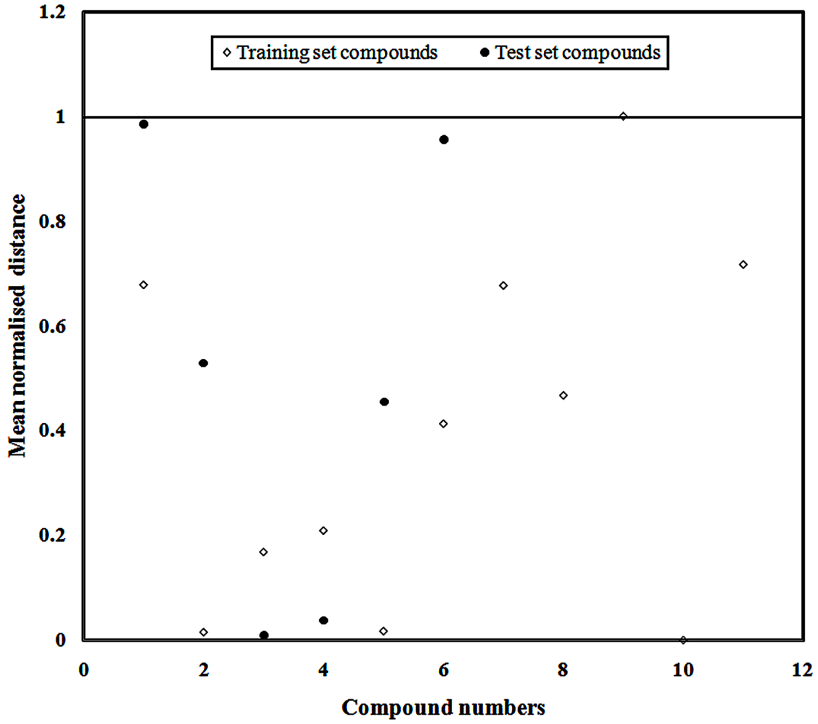


Figure 11. Euclidean distance plot for the arbitrarily developed QSAR model



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

1. A global similarity test should be developed to identify whether the structural features of a new external test compound are enclosed in the training set of compounds from which the QSAR model is developed.
2. There should be a conceptual framework for the assessment of AD for any QSAR model by the regulatory agencies and QSAR experts as no further modification has been done in the OECD guidelines after its establishment.
3. Some effort is needed to investigate the acceptable confidence limits for various AD approaches. Confidence limits could be a useful addition to the AD to predict the new set of compounds more confidently.
4. In order to make the AD estimation more user friendly, QSAR developers have to build up automated software tools where one can easily develop interpolation space as well as scatter plots for respective AD study.
5. There is a need of proper awareness of the AD concept and its implementation in the development of QSAR models. Without a proper AD study, no QSAR model should be accepted by the journals as well as regulatory agencies, as acceptability of any QSAR model is decidedly related with AD.
6. The QSAR model development and the assessment of AD (regarding algorithm and interpretation) should be transparent and it is the major responsibility of the model developer and the regulatory agency to make it clear to the readers and QSAR learners.

CONCLUSION

Applicability domain of a QSAR model is estimated by determining interpolation regions as defined by the training data set in model descriptor space but the region varies depending on the implemented approach. It is interesting to point out that all the implemented AD approaches have their own potential and weakness. Few methods have complex algorithms behind the identification of interpolation regions, on the contrary some have strong and easy statistical background behind the AD estimation. Therefore, it is totally dependent on the model developer how he/she chooses the specific approach according to his/her need to define the applicability domain for the model more accurately and in robust way.

The basic principle of any AD approach is to find out the number of test set compounds which are falling outside of the defined domain of applicability. It is important to note that if someone uses different AD approaches for developed QSAR model from a single dataset, the result may vary in a big way and none of these established approaches can be considered universal one. Therefore, it is always recommended to assess the results from different available possible strategies before assessing a new compound set to get the result with full confidence.

Distance-based approaches are largely used over the other approaches by QSAR developers in recent times. In case of other approaches, the number of test compounds outside the AD is much higher as they are concentrated towards the training set extremities. The distance from training space is significant in defining the model's AD for test set compounds. Again, bad predictions for test compounds are also considered as outside the AD with most of the approaches. The thresholds values are also not fixed for many approaches which may lead to vary the number of compounds outside the AD. So, these issues should be considered more carefully by the QSAR fraternity and regulatory bodies to make the AD studies more transparent and robust ones.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

The following points should be considered by any QSAR model developer while considering AD studies to make his/her QSAR model more robust, predictive and reproducible:

1. The utmost important aspect of any QSAR model is its prediction capability for new set of compounds. The AD study has emerged as one of the key and compulsory validation criteria considering the reliability of predictive capability of QSAR model for a new set of compounds.
2. Considering the basic principle of AD approaches, one has to rely on the AD of the QSAR before estimating activity/property/toxicity of any chemical or pharmaceuticals by the developed QSAR model.
3. The core of AD for any specific QSAR model is the training space which is constructed by the training set of compounds. Therefore, the selection of the training set is very crucial aspect as it should consist of the characteristics of the total dataset to reflect the effect of each compound in the developed model.
4. The response value or endpoint of the QSAR model is determined experimentally. So, the reliability of these values (experiments should be performed in same condition for same endpoints) should be thoroughly checked before constructing the QSAR models. As AD is also dependent on the developed QSAR model, error in response value will mislead the QSAR model which will lead to erroneous interpolation space for AD.
5. There is always a chance of uncertainty related with the assessment of AD for any QSAR model. If the constructed QSAR model is not reliable, one cannot get confidence in AD assessment. Again, mechanistic aspects are rarely considered in the model development by the modeler to make the fact more uncertain one.
6. Estimation of AD by various approaches is highly dependent on a number of issues like the model dimensionality, used descriptors, response value as endpoints, underlying data distribution and last but not the least algorithm of AD determination. So, there is always a chance of large number of assumptions in the AD study.

REFERENCES

Bender, A., & Glen, R. C. (2004). Molecular similarity: A key technique in molecular informatics. *Journal of Organic and Biomolecular Chemistry*, *2*(22), 3204–3218. doi:10.1039/b409813g PMID:15534697

Benigni, R., & Giuliani, A. (2003). Putting the predictive toxicology challenge into perspective: Reflections on the results. *Bioinformatics (Oxford, England)*, *19*(10), 1194–1200. doi:10.1093/bioinformatics/btg099 PMID:12835261

Carlsen, L., Kenessov, B. N., & Batyrbekova, S. Y. (2009). A QSAR/QSTR study on the human health impact of the rocket fuel 1, 1-dimethyl hydrazine and its transformation products: Multicriteria hazard ranking based on partial order methodologies. *Environmental Toxicology and Pharmacology*, *27*(3), 415–423. doi:10.1016/j.etap.2009.01.005 PMID:21783973

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chen, J. J., Tsai, C. A., Young, J. F., & Kodell, R. L. (2005). Classification ensembles for unbalanced class sizes in predictive toxicology. *SAR and QSAR in Environmental Research*, *16*(6), 517–529. doi:10.1080/10659360500468468 PMID:16428129

Dimitrov, S., Dimitrova, G., Pavlov, T., Dimitrova, N., Patlewicz, G., Niemela, J., & Mekenyan, O. A. (2005). Stepwise approach for defining the applicability domain of SAR and QSAR models. *Journal of Chemical Information and Modeling*, *45*(4), 839–849. doi:10.1021/ci0500381 PMID:16045276

Eriksson, L., Jaworska, J., Worth, A. P., Cronin, M. T., 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

Fechner, N., Jahn, A., Hinselmann, G., & Zell, A. (2009). Atomic local neighborhood flexibility incorporation into a structured similarity measure for QSAR. *Journal of Chemical Information and Modeling*, *49*(3), 549–560. doi:10.1021/ci800329r PMID:19434895

Ferreira, M. M. C. (2001). Polycyclic aromatic hydrocarbons: A QSPR study. *Chemosphere*, *44*(2), 125–146. doi:10.1016/S0045-6535(00)00275-7 PMID:11444294

Forina, M., Armanino, C., Leardi, R., & Drava, G. (1991). A class-modelling technique based on potential functions. *Journal of Chemometrics*, *5*(5), 435–453. doi:10.1002/cem.1180050504

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

Gramatica, P. (2007). Principles of QSAR models validation: Internal and external. *QSAR & Combinatorial Science*, *26*(6), 694–701. doi:10.1002/qsar.200610151

Guha, R., & Jurs, P. C. (2005). Determining the validity of a QSAR model-a classification approach. *Journal of Chemical Information and Modeling*, *45*(1), 65–73. doi:10.1021/ci0497511 PMID:15667130

Hair, J. F. Jr, Anderson, R. E., Tatham, R. L., & Black, W. C. (2005). *Multivariate data analysis*. Singapore: Pearson Education.

Jaworska, J., Nikolova-Jeliazkova, N., & Aldenberg, T. (2005). QSAR applicability domain estimation by projection of the training set descriptor space: A review. *Alternatives to Laboratory Animals*, *33*(5), 445–459. PMID:16268757

Jaworska, J. S., Comber, M., Auer, C., & Van Leeuwen, C. J. (2003). Summary of a workshop on regulatory acceptance of (Q)SARs for human health and environmental endpoints. *Environmental Health Perspectives*, *111*(10), 1358–1360. doi:10.1289/ehp.5757 PMID:12896859

Jaworska, J. S., Nikolova-Jeliazkova, N., & Aldenberg, T. (2004). *Review of methods for applicability domain estimation. Report*. Ispra, Italy: The European Commission-Joint Research Centre.

Jouan-Rimbaud, D., Bouveresse, E., Massart, D. L., & de Noord, O. E. (1999). Detection of prediction outliers and inliers in multivariate calibration. *Analytica Chimica Acta*, *388*(3), 283–301. doi:10.1016/S0003-2670(98)00626-6

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Kar, S., & Roy, K. (2010). Predictive toxicology using QSAR: A perspective. *Journal of the Indian Chemical Society*, *87*, 1455–1515.

Manallack, D. T., Tehan, B. G., Gancia, E., Hudson, B. D., Ford, M. G., & Livingstone, D. J. et al. (2003). A consensus neural network-based technique for discriminating soluble and poorly soluble compounds. *Journal of Chemical Information and Computer Sciences*, *43*(2), 674–679. doi:10.1021/ci0202741 PMID:12653537

Mirkin, B. (2005). *Clustering for data mining: A data recovery approach*. London: Chapman & Hall/CRC. doi:10.1201/9781420034912

Netzeva, T. I., Worth, A. P., Aldenberg, T., Benigni, R., Cronin, M. T. D., & Gramatica, P. et al. (2005). Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships. *Alternatives to Laboratory Animals*, *33*(2), 155–173. PMID:16180989

Nikolova, N., & Jaworska, J. (2004). Approaches to measure chemical similarity-a review. *QSAR & Combinatorial Science*, *22*(9-10), 1006–1026.

Nikolova-Jeliazkova, N., & Jaworska, J. (2005). An approach to determining applicability domain for QSAR group contribution models: An analysis of SRC KOWWIN. *Alternatives to Laboratory Animals*, *33*(5), 461–470. PMID:16268758

OECD. (2004). *Principles for the validation of (Q)SARs*. Retrieved 31 March, 2014 from http://www.oecd.org/dataoecd/33/37/37849783.pdf

OECD. (2007). *Guidance document on the validation of (quantitative) 1226 tructure-activity relationships (Q)SARs] models*. ENV/JM/MONO(2007)2.

Perkins, R., Fang, H., Tong, W., & Welsh, W. J. (2003). Quantitative structure-activity relationship methods: Perspectives on drug discovery and toxicology. *Environmental Toxicology and Chemistry*, *22*(8), 1666–1679. doi:10.1897/01-171 PMID:12924569

Preparata, F. P., & Shamos, M. I. (1991). *Computational geometry: An introduction* (F. P. Preparata & M. I. Shamos, Eds.). New York: Springer-Verlag.

Seber, G. A. F. (1984). *Multivariate observations*. New York: John Wiley & Sons. doi:10.1002/9780470316641

Selassie, C. D. (2003). Article. In D. J. Abraham (Ed.), Burger's medicinal chemistry and drug discovery (pp. 1–48). New York: John Wiley.

Sheridan, R., Feuston, R. P., Maiorov, V. N., & Kearsley, S. (2004). Similarity to molecules in the training set is a good discriminator for prediction accuracy in QSAR. *Journal of Chemical Information and Computer Sciences*, *44*(6), 1912–1928. doi:10.1021/ci049782w PMID:15554660

SIMCA-P 10.0. (2002). Retrieved from www.umetrics.com

Smellie, A. (2004). Accelerated K-means clustering in metric spaces. *Journal of Chemical Information and Computer Sciences*, *44*(6), 1929–1935. doi:10.1021/ci0499222 PMID:15554661

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Stanforth, R. W., Kolossov, E., & Mirkin, B. (2007). A measure of domain of applicability for QSAR modeling based on intelligent K-means clustering. *QSAR & Combinatorial Science*, *26*(7), 837–844. doi:10.1002/qsar.200630086

Tetko, I. V. (2008). Associative neural network. *Methods in Molecular Biology (Clifton, N.J.)*, *458*, 185–202. doi:10.1007/978-1-60327-101-1_10 PMID:19065811

Tetko, I. V., Bruneau, P., Mewes, H. W., Rohrer, D. C., & Poda, G. I. (2006). Can we estimate the accuracy of ADME-Tox predictions? *Drug Discovery Today*, *11*(15-16), 700–707. doi:10.1016/j.drudis.2006.06.013 PMID:16846797

Tetko, I. V., Sushko, I., Pandey, A. K., Zhu, H., Tropsha, A., & Papa, E. et al. (2008). Critical assessment of QSAR models of environmental toxicity against Tetrahymena pyriformis: Focusing on applicability domain and overfitting by variable selection. *Journal of Chemical Information and Modeling*, *48*(9), 1733–1746. doi:10.1021/ci800151m PMID:18729318

Tetko, I. V., & Tanchuk, V. Y. (2002). Application of associative neural networks for prediction of lipophilicity in ALOGPS 2.1 program. *Journal of Chemical Information and Computer Sciences*, *42*(5), 1136–1145. doi:10.1021/ci025515j PMID:12377001

Tong, W., Hong, H., Fang, H., Xie, Q., & Perkins, R. (2003). Decision forest: Combining the predictions of multiple independent decision tree models. *Journal of Chemical Information and Computer Sciences*, *43*(2), 525–531. doi:10.1021/ci020058s PMID:12653517

Tong, W., Hong, H., Xie, Q., Xie, L., Fang, H., & Perkins, R. (2004). Assessing QSAR limitations - A regulatory perspective. *Current Computer-aided Drug Design*, *1*(2), 195–205. doi:10.2174/1573409053585663

Tong, W., Xie, Q., Hong, H., Shi, L., Fang, H., & Perkins, R. (2004). Assessment of prediction confidence and domain extrapolation of two structure–activity relationship models for predicting estrogen receptor binding activity. *Environmental Health Perspectives*, *112*(12), 1249–1254. doi:10.1289/ehp.7125 PMID:15345371

TOPKAT OPS. (2000). *U.S. Patent 6, 036, 349, March 14, 2000*. Washington, DC: US Patent Office.

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

Walker, J. D., Jaworska, J., Comber, J. 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

Weaver, S., & Paul Gleeson, M. (2008). The importance of the domain of applicability in QSAR modeling. *Journal of Molecular Graphics & Modelling*, *26*(8), 1315–1326. doi:10.1016/j.jmgm.2008.01.002 PMID:18328754

Wold, S., Sjostrom, M., & Eriksson, L. (2001). PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, *58*(2), 109–130. doi:10.1016/S0169-7439(01)00155-1

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Worth, A. P., Bassan, A., Gallegos, A., Netzeva, T. I., Patlewicz, G., & Pavan, M. et al. (2005). *The characterisation of (quantitative) structure-activity relationships: preliminary guidance.* Ispra, Italy: European Commission, Joint Research Centre.

ADDITIONAL READING

Atkinson, A. C. (1985). *Plots, Transformations and Regression*. Oxford: Clarendon Press.

Baskin, I. I., Kireeva, N., & Varnek, A. (2010). The one-class classification approach to data description and to models applicability domain. *Molecular Informatics*, *29*(8-9), 581–587. doi:10.1002/minf.201000063

berg, T. (2004). A QSAR for baseline toxicity: validation, domain of application, and prediction. *Chemical Research in Toxicology, 17*(12), 1630-1637.

Blekhman, I. I., & Yaroshevich, N. P. (2004). Extension of the domain of applicability of the integral stability criterion (extremum property) in synchronization problems. *Journal of Applied Mathematics and Mechanics*, *68*(6), 839–846. doi:10.1016/j.jappmathmech.2004.11.005

Cronin, M. T. D., & Schultz, T. W. (2003). Pitfalls in QSAR. *Journal of Molecular Structure THEOCHEM*, *622*(1-2), 39–51. doi:10.1016/S0166-1280(02)00616-4

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

Dragos, H., Gilles, M., & Varnek, A. (2009). Predicting the predictability: A unified approach to the applicability domain problem of QSAR models. *Journal of Chemical Information and Modeling*, *49*(7), 1762–1776. doi:10.1021/ci9000579 PMID:19530661

Ellison, C. M., Sherhod, R., Cronin, M. T. D., Enoch, S. J., Madden, J. C., & Judson, P. N. (2011). Assessment of methods to define the applicability domain of structural alert models. *Journal of Chemical Information and Modeling*, *51*(5), 975–985. PMID:21488656

Eriksson, L., Jaworska, J., Worth, A., 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), 1351–1375. doi:10.1289/ehp.5758 PMID:12896860

Furusj, E., Svenson, A., Rahmberg, M., & Andersson, M. (2006). The importance of outlier detection and training set selection for reliable environmental QSAR predictions. *Chemosphere, 63*(1), 99-108.

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

Kolossov, E., & Stanforth, R. (2007). The quality of QSAR models: Problems and solutions. *SAR and QSAR in Environmental Research*, *18*(1), 89–100. doi:10.1080/10629360601053984 PMID:17365961

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Kubinyi, H. (1998). Similarity and dissimilarity: A medicinal chemist's view. *Perspectives in Drug Discovery and Design*, *9-11*(0), 225–252. doi:10.1023/A:1027221424359

Kuhne, R., Ebert, R.-E., & Schuurman, G. (2009). Chemical domain of QSAR models from atom-centered fragments. *Journal of Chemical Information and Modeling*, *49*(12), 2660–2669. doi:10.1021/ci900313u PMID:19928752

Leonard, J. T., & Roy, K. (2006). On selection of training and test sets for the development of predictive QSAR models. *QSAR & Combinatorial Science*, *25*(3), 235–251. doi:10.1002/qsar.200510161

Sahlin, U., Jeliazkova, N., & Öberg, T. (2014). Applicability Domain Dependent Predictive Uncertainty in QSAR Regressions. *Molecular Informatics*, *33*(1), 26–35. doi:10.1002/minf.201200131

Schroeter, T. S., Schwaighofer, A., Mika, S., Laak, A. T., Suelzle, D., & Ganzer, U. et al. (2007). Estimating the domain of applicability for machine learning QSAR models: A study on aqueous solubility of drug discovery molecules. *Journal of Computer-Aided Molecular Design*, *21*(9), 485–498. doi:10.1007/s10822-007-9125-z PMID:17632688

Schultz, T. W., Hewitt, M., Netzeva, T. I., & Cronin, M. T. D. (2007). Assessing Applicability Domains of Toxicological QSARs: Definition, Confidence in Predicted Values, and the Role of Mechanisms of Action. *QSAR & Combinatorial Science*, *26*(2), 238–254. doi:10.1002/qsar.200630020

Schultz, T. W., Netzeva, T. I., & Cronin, M. T. (2004). Evaluation of QSARs for ecotoxicity: A method for assigning quality and confidence. *SAR and QSAR in Environmental Research*, *15*(5-6), 385–397. doi:10.1080/10629360412331297344 PMID:15669697

Sheridan, R. P. (2012). Three Useful Dimensions for Domain Applicability in QSAR Models Using Random Forest. *Journal of Chemical Information and Modeling*, *52*(3), 814–823. doi:10.1021/ci300004n PMID:22385389

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Boca Raton: CRC Press. doi:10.1007/978-1-4899-3324-9

Soto, A. J., Vazquez, G. E., Strickert, M., & Ponzoni, I. (2011). Targetdriven subspace mapping methods and their applicability domain estimation. *Molecular Informatics*, *30*(9), 779–789. doi:10.1002/minf.201100053

Sushko, I., Novotarskyi, S., Körner, R., Pandey, A. K., Cherkasov, A., & Li, J. et al. (2010). Applicability domain for classification problems: Benchmarking of distance to models for Ames mutagenicity set. *Journal of Chemical Information and Modeling*, *50*(12), 2094–2111. doi:10.1021/ci100253r PMID:21033656

Sushko, I., Novotarskyi, S., Körner, R., Pandey, A. K., Kovalishyn, V. V., Prokopenko, V. V., & Tetko, I. V. (2010). Applicability domain for in silico models to achieve accuracy of experimental measurements. *Journal of Chemometrics*, *24*(3-4), 202–208. doi:10.1002/cem.1296

Todeschini, R., Consonni, V., & Pavan, M. (2004). A distance measure between models: A tool for similarity/diversity analysis of model populations. *Chemometrics and Intelligent Laboratory Systems*, *80*(1), 55–61. doi:10.1016/j.chemolab.2003.10.003

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Toropova, A. A., & Benfenati, E. (2008). Additive SMILES-based optimal descriptors in QSAR modelling bee toxicity: Using rare SMILES attributes to define the applicability domain. *Bioorganic & Medicinal Chemistry*, *16*(9), 4801–4809. doi:10.1016/j.bmc.2008.03.048 PMID:18395455

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

Tropsha, A., & Golbraikh, A. (2007). Predictive QSAR modeling workflow, model applicability domains, and virtual screening. *Current Pharmaceutical Design*, *13*(34), 3494–3504. doi:10.2174/138161207782794257 PMID:18220786

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

KEY TERMS AND DEFINITIONS

3Rs: The 3Rs represent the three phrases "Reduction", "Replacement" and "Refinement". Replacement corresponds to the use of non-living materials to substitute conscious living higher animals. Reduction refers to the decrease in number of animals, and Refinement means reduction in the harshness of inhuman methodologies to be applied onto the experimental animals. *In silico* approaches like QSAR is an important alternative tool to support the 3Rs approach.

Applicability Domain: The applicability domain (AD) is a theoretical region explained by physicochemical, structural or biological space and evaluated by the model descriptors and modeled response. In simple explanation, AD is the knowledge or information based on which the training set of the model is constructed, and the model is applicable to make predictions for new compounds within the specific domain.

Multiple Linear Regression: Multiple linear regression (MLR**)** attempts to model the relationship between two or more explanatory variables and a response variable by fitting a linear equation to observed data.

Partial Least Squares: Partial Least Squares (PLS) is a generalization of regression by which strongly correlated and/or noisy or numerous X variables can be handled meticulously. The linear PLS model finds "new variables" (which can be considered as latent variables or X scores), which are linear combinations of the original ones. To avoid the problem of overfitting, a strict test for the significance of each consecutive PLS component is necessary and then stopping when the components are non-significant.

Principal Component Analysis: Principal Component Analysis (PCA) sums up all variation in X into a few more informative new variables called T scores. These new variables are linearly weighted combinations of the original X-variables. The weighting profiles are called loadings, P. For each score variable, the influence (weight) of the original variables is found by looking at its corresponding loading profile.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

QSAR: Quantitative structure-activity relationship (QSAR) is a statistical modelling approach that correlates the molecular structure to a specific activity or property or toxicity derived either *in vitro* or *in vivo*. A simple mathematical relationship in any kind of QSAR method can be demonstrated as the following equation: *Biological activity = f (chemical structure or property)*. QSAR is largely used in drug discovery, environmental fate modeling, property and biological activity prediction of new, untested compounds from the molecular structures of compounds.

Randomization: Randomization is a validation technique which is performed in order to ensure the robustness of the developed QSAR model. The values of the dependent variable (Y) are randomly shuffled and new QSAR models are developed keeping the independent variable matrix unchanged. This method is of two types: process randomization and model randomization performed at varying confidence levels. The process is also known as Y-scrambling or Y-randomization technique.

Standard Deviation: Standard deviation (SD) can be defined as the measure of the deviation or dispersion of a set of data from its mean. SD can be calculated as the square root of variance.

Validation: Validation is one of the backbone steps for *in silico* models, which includes evaluation of issues such as data quality, robustness, predictability of the model and mechanistic interpretability in addition to statistical judgment. Validation strategies are largely dependent on various statistical validation metrics.

Virtual Screening: Virtual screening (VS) is a computational technique used in drug discovery to explore various existing libraries of small drug molecules in order to identify the promising structures which are most likely to bind to a drug target, especially a protein receptor or enzyme.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 6
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: 10.4018/978-1-4666-8136-1.ch006

Copyright © 2015, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/14/2023 7:16 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 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)

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 2/14/2023 7:16 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)

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 determination 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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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.

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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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):

$$\log \frac{1}{C} = a(\log P)^2 + b\,log\,P + cs + \cdots k \tag{9}$$

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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.

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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

- **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 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 nonlinear 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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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).

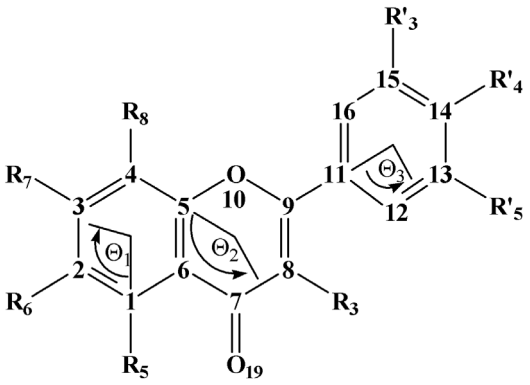
In the QSAR analysis performed by Deeb and Clare (Deeb & Clare, 2007). they have investigated the flavonoid-inhibitory activity of 54 analogues using the nodal angle descriptors (Clare, 2000; Clare & Supuran, 2005) and flipstep regression analysis described elsewhere (Clare, 2000; 2001a). For the flavonoid, 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.

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).

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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.



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 2/14/2023 7:16 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 2/14/2023 7:16 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 2/14/2023 7:16 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 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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:

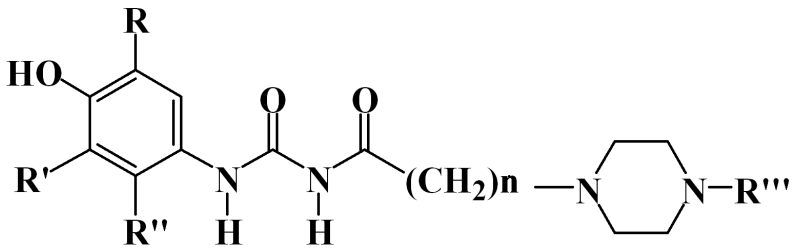
- **MPC:** Most positive charges,
- **Qmean:** The Mean absolute charge (charge polarization),
- **SSC:** Sum of squares of charges,
- **X5:** Connectivity index Chi-5,
- **S0K:** Kier symmetry index,
- **PW2:** Path/walk-2-randic shape index,
- **DMz:** Molecular dipole moment at z-direction,
- **DM_t:** 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

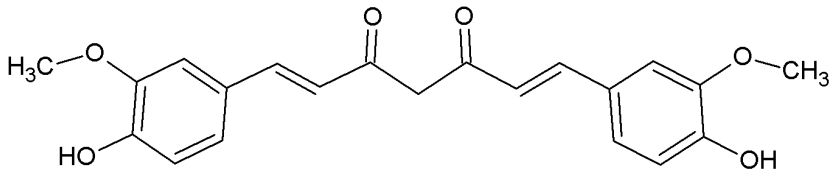
In 2014 (Chen, Zhu, Chen, Dong & Li,) performed 3D-QSAR study on the antioxidant capacity or curcumin analogues. A comparative molecular similarity indices analysis (CoMSIA) was done on a set of 27 curcumin-like diarylpentanoid analogues and their DPPH scavenging activities. This study indicates that 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 correlation 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.

Figure 2. General structure of hydroxyphenylureas



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 3. 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 (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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

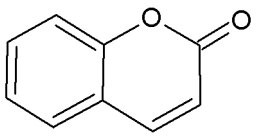
set was divided into training (25 compounds) and test (8 compounds) sets using the *k*-means clustering technique and external validation was done based on the activity prediction of the test set compounds. $R^2_{pred} = 0.917$ (for the test set) indicates significant ability of the developed model to 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 different 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,

Figure 4. Parent coumarin nucleus used in this study



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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. In other words, the models indicate that the antiradical activity increases as we incorporate hydrophilic groups to the coumarin molecules.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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.

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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.

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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).

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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 7
QSAR Studies on Bacterial Efflux Pump Inhibitors

Khac-Minh Thai
University of Medicine and Pharmacy at HCMC, Vietnam

Thuy-Viet-Phuong Nguyen
University of Medicine and Pharmacy at HCMC, Vietnam

Trong-Nhat Do
University of Medicine and Pharmacy at HCMC, Vietnam

Duc-Khanh-Tho. Nguyen
University of Medicine and Pharmacy at HCMC, Vietnam

Thanh-Dao Tran
University of Medicine and Pharmacy at HCMC, Vietnam

ABSTRACT

Antimicrobial drug resistance occurs when bacteria undergo certain modifications to eliminate the effectiveness of drugs, chemicals, or other agents designed to cure infections. To date, the burden of resistance has remained one of the major clinical concerns as it renders prolonged and complicated treatments, thereby increasing the medical costs with lengthier hospital stays. Of complex causes for bacterial resistance, there has been increasing evidence that proved the significant role of efflux pumps in antibiotic resistance. Coadministration of Efflux Pump Inhibitors (EPIs) with antibiotics has been considered one of the promising ways not only to improve the efficacy but also to extend the clinical utility of existing antibiotics. This chapter begins with outlining current knowledge about bacterial efflux pumps and drug designs applied in identification of their modulating compounds. Following, the chapter addresses and provides a discussion on Quantitative Structure-Activity Relationship (QSAR) analyses in search of novel and potent efflux pump inhibitors.

DOI: 10.4018/978-1-4666-8136-1.ch007

Copyright © 2015, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

INTRODUCTION

Antimicrobial drugs have been important tools of healthcare in several decades because of their effectiveness in control of bacterial infections. Unfortunately, soon after their invention it was realized that some pathogens rapidly developed resistance to antibiotics (Neu, 1992; Wood, Gold, & Moellering Jr, 1996). People infected with antimicrobial-resistant organisms are more likely to have longer, more expensive hospital stays, and may be more likely to die as a result of the infection. Initially, this problem was overcome by discovery of new classes of antibiotics such as aminoglycosides, macrolides and glycopeptides, but the bacteria rapidly showed an impressive array of defensive mechanisms that conferred on them, resistance to many modes of attack (Wood et al., 1996). The main mechanisms whereby the bacteria develop resistance to antimicrobial agents include enzymatic inactivation (Bush & Miller, 1998; Sabatini et al., 2012), modification of the antibiotic attack site(s) (Ruiz, 2003; Sabatini et al., 2012), and reduction of intracellular drug concentration by changes in membrane permeability (Nikaido, 2003; Sabatini et al., 2012) or by the overexpression of efflux pumps (Li & Nikaido, 2009; Sabatini et al., 2012).

One primary mechanism of antibiotic resistance is extrusion of the foreign chemical, which is termed efflux. In 1980, tetracycline was reported that it could be actively effluxed from the bacterial cell (McMurry, Petrucci, & Levy, 1980). From then on, many efflux-related mechanisms have been discovered. Overexpression of these efflux pumps may lead to antibiotic resistance. While efflux pump proteins are present in both Gram-positive and Gram-negative bacteria and also in eukaryotes, antibiotic resistance due to efflux is more of a problem in Gram-negative bacteria than in Gram-positive bacteria (Nikaido, 1996). This is because the presence of an outer membrane in Gram-negative bacteria shows comparatively lower permeability and complements the efflux activity of these pumps. Besides, studies of efflux pumps in Gram-negative bacteria have got more complicated as their double-membrane cells allow the expression of a tripartite efflux pump system such as AcrA/AcrB/TolC in Enterobacteriaceae, or MexA/MexB/OprM in *Pseudomonas aeruginosa*. One plausible practice to fight against multidrug efflux systems (MES) is the combination of conventional antimicrobial agents/antibiotics with small molecules that block MES known as multidrug efflux pump inhibitors (EPIs). An array of approaches in academic and industrial research settings, varying from high-throughput screening (HTS) ventures to bioassay guided purification and determination, have yielded a number of promising EPIs in a series of pathogenic systems (Tegos et al., 2011).

Up to now, most inhibitors of efflux pumps have been discovered through traditional random screening of synthetic compounds or natural products libraries. The assays used are very simple and easily adapted to high-throughput screening. An alternative approach is to screen libraries of known drugs. The identification of a novel mode of action in an approved drug could considerably shorten the development route and lessen the risks associated with a new chemical entity.

OVERVIEW OF BACTERIAL EFFLUX PUMPS AND ANTIBIOTIC RESISTANCE

Bacterial Efflux Pumps

Efflux pumps in Gram-positive bacteria can be categorized into four families (Handzlik, Matys, & Kieć-Kononowicz, 2013), namely ABC (ATP-binding cassette) (Higgins, 2001), MFS (major facilitator superfamily) (Saier Jr et al., 1999), SMR (small multidrug resistance) (Chung & Saier Jr, 2001) (Jack, Yang, & H Saier, 2001), and MATE (multidrug and toxic compound extrusion) (Hvorup et al., 2003).

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

For Gram-negative bacteria, many pump systems including *Campylobacter jejuni* (CmeABC) (Pumbwe & Piddock, 2002) (Lin, Michel, & Zhang, 2002), *Escherichia coli* (AcrAB-TolC, AcrEF-TolC, EmrB, EmrD) (Poole, 2000), *Pseudomonas aeruginosa* (MexAB-OprM, MexCD-OprJ, MexEF-OprN and MexXY-OprM) (Poole, 2000), *Salmonella typhimurium* (AcrAB) (Nikaido, 2001) have been described. These pumps basically belong to five major families, including the MFS (major facilitator superfamily), MATE (multidrug and toxic extrusion), SMR (small multi-drug resistance), ABC (ATP-binding cassette) and RND (resistance-nodulation-division) families (Saier Jr, 1998). Co-expression of multiple types of efflux pumps can bring an additive or multiplicative effect on drug resistance (Lee et al., 2000).

MFS is the largest superfamily of transporters involved in symport, uniport, or antiport of various small solutes (Saier Jr et al., 1999). Examples include sugars, neurotransmitters, amino acids, Krebs cycle metabolites, and importantly, drugs. Most MFS efflux pumps have 400-600 amino acid residues and possess either 12 or 14 putative transmembrane domains. In MFS transporters, negatively charged amino acid residues located in transmembrane helices appear to play a critical role in the protonation/deprotonation step during transport. Typical MFS efflux pumps in Gram-positive bacteria include NorA, NorB, MdeA, Tet38 (*Staphylococcus aureus*), LmrB, Bmr, Bmr3, Blt (*Bacillus subtilis*), MefA (*Streptococcus pyogenes*), MefE (*Streptococcus pneumoniae*) or CmlR (*Streptococcus coelicor*) (Borges-Walmsley, McKeegan, & Walmsley, 2003; Handzlik et al., 2013; Jarmuła, Obłąk, Wawrzycka, & Gutowicz, 2010; Markham & Neyfakh, 2001; Paulsen, Brown, & Skurray, 1996; Poole, 2005).

The SMR family consists of small multidrug transporters widespread among eubacteria (Bay, Rommens, & Turner, 2008). These proteins are about 100-amino acid residues long, with four transmembrane helices. The examples of SMR efflux pumps in Gram-positive bacteria are EbrAB (*Bacillus subtilis*) or Smr, QacG, QacH (*Staphylococcus aureus*) (Jarmuła et al., 2010; Kikukawa, Nara, Araiso, Miyauchi, & Kamo, 2006; Putman, van Veen, & Konings, 2000).

All ABC transporters contain four essential modules, two nucleotide binding domains (NBDs) and two transmembrane domains (TMDs). These four modules can be encoded by four separate genes or fused pairwise in all possible combinations (Hollenstein, Dawson, & Locher, 2007; Lubelski, Konings, & Driessen, 2007). Bacterial MultiDrug Resistance (MDR) is usually homo- or heterodimers in which one NBD is fused to one TMD. LmrA (*Lactococcus lactis*) and Rv1217c-Rv1218c (*Mycobacterium tuberculosis*) (Borges-Walmsley et al., 2003; Jarmuła et al., 2010; Wang et al., 2013) belong to this transport protein family.

The RND transporters are responsible for the high intrinsic antibiotic resistance which can be seen in Gram-negative bacteria, one component of the notorious "natural superbug" phenotype. They are also found in Gram-positive bacteria where their functions are largely unknown (Tseng et al., 1999). RND transporters are homo- or heterotrimers consisting of promoters of about 1100-amino acid residues in size. Each promoter consists of the 12 transmembrance helices and a large periplasmic domain.

Most MATE transporters consist of 400-550 amino acid residue polypeptides with 12 putative transmembrane helices which share about 40% sequence similarity. An example of MATE efflux pump in Gram-positive bacteria is MepA protein found in *Staphylococcus aureus* (Handzlik et al., 2013; Omote, Hiasa, Matsumoto, Otsuka, & Moriyama, 2006; Wasaznik, Grinholc, & Bielawski, 2008). Currently, no high-resolution structures are available for this family of transporters. Secondary structure predictions showed the symmetric repetition of conserved regions in the *N*- and *C*-terminal halves of the proteins suggesting a gene duplication event.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Bacterial Efflux Pump Inhibitors

For Gram-positive bacteria, the search for new EPIs performed in the last decade was mostly focused on the NorA of *S. aureus* (Handzlik et al., 2013). Many approaches have been made to explore new EPIs. EPI activity is often found by antibiotic–EPI combination assays using efflux-proficient and -deficient *S. aureus* strain pairs. Efflux inhibition is verified with fluorometric efflux assays using ethidium bromide, a nonspecific DNA intercalator. A number of EPIs targeting *S. aureus* NorA have been identified including current chemotherapeutic agents used as EPIs (e.g. reserpine, P-gp inhibitors, verapamil, proton pump inhibitors, phenothiazines, thioxanthenes, paroxetine and COX-2 inhibitors), naturally occuring EPIs (e.g. 4',5'-*O*-dicaffeoylquinic acid, *N-trans*-feruloyl 4'-*O*-methyldopamine, resin glycosides, pheophorbide A, diterpenes, flavonoids and capsaicin) and synthetic EPIs (e.g. piperine analogs, fluoroquinolone analogues, indole-based inhibitors, berberine-INF55 hybrids, substituted dihydronaphthalene and mesolonic compounds); of which hybrid molecules were suggested for further investigation. An EPIs must be nontoxic and highly selective in order to be a potential candidate for clinical use. In this context, reserpine cannot enter further developments due to its neurotoxicity at concentrations required to inhibit NorA (Markham & Neyfakh, 1996; Schindler, Jacinto, & Kaatz, 2013; Zhang & Ma, 2010).

For Gram-negative bacteria, promethazine, a phenothiazine, combines with penicillin G have been used to inhibit RND pumps in *E. coli* (Lehtinen & Lilius, 2007), as well as inhibit ABC pumps in yeast (Kolaczkowski, Michalak, & Motohashi, 2003). Phenothiazines are able to reverse MDR phenotypes of pathogenic bacteria like *P. aeruginosa* or *S. typhimurium* (Michalak, Wesolowska, Motohashi, Molnar, & Hendrich, 2006; Molnar et al., 1996). Paroxetine can inhibit AcrAB-TolC pump of *E. coli* (German, Kaatz, & Kerns, 2008; Kaatz, Moudgal, Seo, Hansen, & Kristiansen, 2003; Munoz-Bellido, Munoz-Criado, & Garcìa-Rodrìguez, 2000). It is pointed out that arylpiperazines reverse MDR in bacteria overexpressing AcrAB and AcrEF pumps. Moreover, NMP (1-naphthylmethylpiperazine) is the most effective inhibitor that enhances the intracellular concentration of drugs like chloramphenicol, tetracycline, linezolid, macrolides and fluoroquinolones (Kern et al., 2006). The structures with dihalogens among arylpiperidines have been shown to restore the antimicrobial activity of linezolid in *E. coli* (Thorarensen et al., 2001).

Quinolines have been shown the ability to re-establish various antibiotic activities like quinolones, cyclines and chloramphenicol. They are now applied as broad spectrum inhibitors for resistant *E. aerogenes* and *K. pneumoniae* to make them susceptible for the antimicrobials (Chevalier et al., 2004). The analogs of aminoglycoside paromomycin have been applied as inhibitors of the efflux pumps using the bacterium *H. influenzae* (Alekshun & Nelson, 2004). AcrAB efflux pump was inhibited by using antibodies in *E. coli* and this strategy has been patented (Oethinger & Levy, 2002). A recent report has introduced the use of antisense phosphorothioate oligonucleotide encapsulated in a novel anion liposome to restore the activity of fluoroquinolones in *E. coli* (Meng et al., 2011). This method could be utilized for a wide variety of pumps with known gene sequences. Using antibody or parts thereof in order to inhibit MexAB-OprM pumps in *P. aeruginosa* has been patented (Inoko & Yoshihara, 2011).

In addition, a list of about 200,000 synthetic and natural compounds was tested with the aim of finding potential compounds preventing the activity of levofloxacin against *P. aeruginosa* (Chamberland et al., 2002; Coban, Ekinci, & Durupinar, 2004; Lomovskaya et al., 2001; Lomovskaya & Watkins, 2001; Renau et al., 1999). MC-207, 110 (Phenylalanine Arginyl β - Naphthylamide/PAβN) was identified as an inhibitor of Mex pumps. The crystal structure of *E. coli* AcrB has been solved in the presence of MC-207 and its binding to the pump has been proved (Edward, Aires, McDermott, & Nikaido, 2005). This compound has been shown to decrease the frequency of the emergence of highly levofloxacin resistant

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

P. aeruginosa strains and also to reduce the intrinsic resistance of the bug to levofloxacin 8-folds (Coban et al., 2004; Lomovskaya et al., 2001; Renau et al., 1999). It is a competitive inhibitor of the efflux pumps and operates by binding to the same pocket or at a site closer to the antibiotic substrate binding site (Mahamoud, Chevalier, Alibert-Franco, Kern, & Pagès, 2007; Pagès, Masi, & Barbe, 2005). The compound has not only restored the activity of levofloxacin but has also been found to potentiate the activity of other antibiotics such as oxazolidinones, chloramphenicol, rifampicin, macrolides/ ketolides (Zechini & Versace, 2009). Furthermore, it has these effects not only *for P. aeruginosa*, but also for *K. pneumoniae*, *C. jejuni*, *E. coli*, *S. typhimurium* and *E. aerogenes* (Malléa, Chevalier, Eyraud, & Pagès, 2002; Mazzariol, Tokue, Kanegawa, Cornaglia, & Nikaido, 2000).

Drug Design Targets Bacterial Efflux Pumps

Structure-based drug design plays an important part of the success story in the discovery of new drug leads. It is known that a thorough understanding of MDR transporter structure and function, as well as mechanism, would considerably facilitate the drug discovery. There have been, however, the scarcity of atomic structures of these membrane proteins due to difficulties related to the process of protein expression and crystallization (Chang, Ray, & Swaan, 2005; Rosenbusch, 2001). Additionally, it was suggested that there are some limitations associated with protein structures solved by crystallography that should be taken into consideration (DePristo, de Bakker, & Blundell, 2004). To date, the number of available bacterial efflux pump structures is still small compared to the number of identified bacterial transporters, which were described in previous reviews (Delmar, Su, & Yu, 2014; S. Kumar & Varela, 2012; Li & Nikaido, 2009; Van Bambeke & Lee, 2006; Zechini & Versace, 2009). This has hindered the computer-aided discovery of efflux pump chemosensitizers. The Table 1 enumerates MDR transporters with determined structures, of which recently structure-found transporters have mostly belonged to *Escherichia coli* and *Pseudomonas aeruginosa*.

Regarding MDR Gram-negative bacteria, RND transporter AcrAB-TolC and its homologues (including MexAB-OprM and MexXY-OprM) are main contributors to bacterial survival during infection. Certain studies on *Pseudomonas aeruginosa* MexAB-OprM pump, using both ligand-based and structure-based pharmacophore approaches in combination with docking method, recommended several important structural modification favoring inhibitory potentiation of known inhibitors: the incorporation of a hydrophobic group at the 2-position of the pyridopyrimidine scaffold or the inclusion of an olefin spacer between the tetrazole and pyridopyrimidine scaffold can render a boost in potency of this derivative, hydrophilic substitution was feasible on a piperidine moiety at the 2-position without adversely affecting activity, paving the way for the increase of aqueous solubility (Nakayama et al., 2004); in another study, the compound ASN05108137 exhibited maximum similarity to the pharmacophore hypothesis of MC207110, the first inhibitor against MexAB-OprM efflux system of *P.aeruginosa*, and also showed the druglikeness property satisfying the Lipinski's rule of five (Aparna, Mohanalakshmi, Dineshkumar, & Hopper, 2014).

In a recent study, the crystal structure of AcrB and MexB bound to the pyridopyrimidine derivative ABI-PP was determined. This result provides detailed insight into the mechanism by which the inhibitor ABI-PP acts on RND transporters thereby generating a structural basis for designing a universal inhibitor of multidrug efflux transporters (Nakashima et al., 2013). A pharmacophore study was aimed to circumvent the drug resistance in *P.aeruginosa* and *E.coli* by identifying natural compounds from plants having effect on MexB and its counterpart AcrB. Lanatoside C and daidzein were found to have a high possibility to counteract the MDR pumps in the resistance bacteria (Aparna, Dineshkumar, Mohanalakshmi, Velmurugan, & Hopper, 2014).

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 1. Bacterial MDR efflux pumps with determined X-ray structures

Bacterial Strain	Protein Family	Efflux Pump	PDB Code	Resolution (Å)	Bound Ligand	Reference and Year of Release
Bacillus subtilis	MFS	BmrR	3Q5P	2.94	tetracycline	(Bachas, Eginton, Gunio, & Wade, 2011)
			3Q5R	3.05	kanamycin	(Bachas et al., 2011)
			3Q5S	3.10	acetylcholine	(Bachas et al., 2011)
Aquifex aeolicus	MFS	EmrA	4TKO	2.85	-	(Hinchliffe et al., 2014)
Neisseria gonorrhoeae	RND	MtrD	4MT1	3.54	-	(H.-T. Lei et al., 2014)
Campylobacter jejuni	RND	CmeR	3QQA	2.20	taurocholic acid	(H. T. Lei et al., 2011)
			3QPS	2.35	cholic acid	(H. T. Lei et al., 2011)
Vibrio cholerae	MATE	NorM	3MKT	3.65	-	(He et al., 2010)
			3MKU	4.20	rubidium ion	(He et al., 2010)
Escherichia coli	ABC	MacA	3FPP	2.99	-	(Yum et al., 2009)
	RND	AcrB	4CDI	3.70	-	(Du et al., 2014)
			4C48	3.30	dodecyl-β-D-maltoside and nickel (II) ion	(Du et al., 2014)
			3W9H	3.05	ABI-PP	(Nakashima et al., 2013)
			2HRT	3.00	citrate anion	(Seeger et al., 2006)
		TolC	2VDE	3.20	chloride ion	(Bavro et al., 2008)
			1EK9	2.10	-	(Koronakis, Sharff, Koronakis, Luisi, & Hughes, 2000)
Pseudomonas aeruginosa	RND	MexB	3W9I	2.71	dodecyl-β-D-maltoside	(Nakashima et al., 2013)
			3W9J	3.15	dodecyl-β-D-maltoside and ABI-PP	(Nakashima et al., 2013)
			2V50	3.00	dodecyl-β-D-maltoside	(Sennhauser, Bukowska, Briand, & Grütter, 2009)
		OprM	3D5K	2.40	chloride ion and sodium ion	(Phan et al., 2010)
			1WP1	2.56	-	(Akama et al., 2004)

(From Protein Data Bank).

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Another pharmacophore hypothesis was generated by the HipHop method to analyze the structure-activity relationships of active sites of protein-ligand interactions thereby increasing the knowledge of structure and mechanism of RND-type AdeABC EPIs in *Acinetobacter baumannii*. The achieved 3D common feature pharmacophore model revealed that two hydrogen bond acceptors (HBA) and three hydrophobic aromatic (HpAr) were found to be significant for binding to the active site of the target protein. Three HpAr features displayed the required substitution of bulky aromatic moieties, while two HBA atoms or groups were necessary in the molecule to bind to the target protein (Yilmaz et al., 2014).

More recently, structures of interest for rational design include the advent of the pseudo-atomic structure of complete AcrAB-TolC in complex with AcrZ, a modulatory protein partner from *E.coli*. It was found that the studied structure is likely to be similar to homologous assemblies in pathogenic species that infect humans such as *Vibrio cholerae, P.aeruginosa, Neisseria gonorrhoeae* and *Salmonella enterica* (Du et al., 2014).

QSAR APPROACHES IN SEARCH FOR EFFLUX PUMP INHIBITORS

QSAR Approaches in Search for EPIs in Gram Positive Bacteria

The methicillin-resistant *Staphylococcus aureus* (MRSA) is one of the most frequent nosocomial pathogens in developed countries. Among the Gram positive bacteria, *S.aureus* is responsible for a large number of deaths worldwide. MRSA is resistant to a variety of antibiotics including tetracyclines, aminoglycosides and fluoroquinolones. Studies found that NorA is a predominant protein efflux pump responsible for efflux mechanisms in *S.aureus* (Sabatini et al., 2012). For these reasons, NorA has been one of the most studied Gram-positive bacterial efflux pumps in the search for efflux pump inhibitors. Although other protein targets have been investigated and there are some newly found derivatives proved to have inhibitory activities towards those proteins, there is still a lack of QSAR studies focused on such inhibitors. With respect to NorA in *S.aureus*, while there have been a number of natural and synthetic product known to be inhibitors of NorA efflux pump, there is still a need to discover new and potent NorA inhibitors. However, the slow emergence of high-resolution structure of this pump has hampered the intelligent design of its modulators, necessitating a ligand-based approach. With currently available structural data regarding known NorA inhibitors, robust QSAR models were generated, paving the way for preliminary identification of compounds with NorA inhibitory activities, whereby more thorough research can be conducted in order to identify the inhibitors with high potentiality to become future "adjuvants" of antibiotics.

QSAR of Piperine Analogs for Bacterial NorA Efflux Inhibitors

Piperine, a major component of Piper nigrum, is a putative bacterial EPI (Khan, Mirza, Kumar, Verma, & Qazi, 2006). In addition, piperine has been found to be associated with several activities (Amit Nargotra et al., 2009): inhibiting several cytochrome P450-mediated pathways and phase II reactions in animal models (Atal, Zutshi, & Rao, 1981; Singh, Dubey, & Atal, 1986); inhibiting human P-glycoprotein (Bhardwaj et al., 2002) and potentially inhibiting rat hepatic microsomal constitutive and inducible cytochrome P450 activities (Koul et al., 2000).

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

After screening a chemical library consisting of 200 structurally diverse compounds for inhibition of the NorA efflux pump, three piperine analogues, namely SK-20, SK-56 and SK-29, were found to be the most potent inhibitors of the NorA efflux pump. The structural adjustments of piperine that can lead to bacterial efflux pump inhibitors with significantly higher activities include: (i) the introduction of an alkyl group at the C-4 position mainly contributes to potentiation, and ethyl and n-propyl show highest effect; (ii) substitution of a piperidinyl moiety by an aromatic amine such as anisidine or toluidine renders greatest potentiation, whereas other substituents such as aniline, amino esters, pyrollidine, azepine and alkylamines (except isobutyl amine) appear to be less effective; (iii) unsaturation is a critical feature for potentiation, as di- and tetra-hydro piperine show much less activity than corresponding unsaturated analogues; (iv) the omission of amide carbonyl group leads to a decrease in potentiation; and (v) the presence of a 3,4-methylenedioxyphenyl or 4-methoxyphenyl group in structure of piperine derivatives can enhance potentiation (A. Kumar et al., 2008).

In the QSAR study of piperine analogs for bacterial NorA efflux pump inhibitors (Amit Nargotra et al., 2009), researchers at Indian Institute of Integrative Medicine, India employed the Cerius2 4.10 software package (from Accelrys Inc.) to calculate 2D descriptors differentiating the molecules in terms of their size, degree of branching, flexibility, and overall shape. A total set of 25 compounds was divided into 20 molecules in training set and 5 molecules in test set. The general structures of the compounds are given in Figure 1.

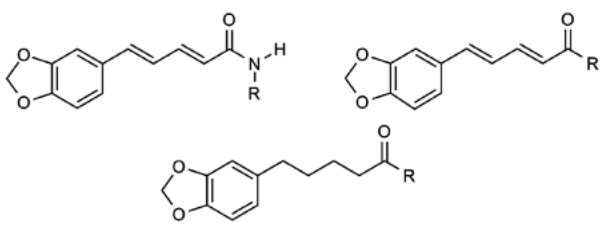
All the molecules were energy minimized using the Cerius2 OFF module with default parameters of Smart Minimizer. The QSAR model was generated using the genetic function approximation (GFA) method and its statistical significance was validated using randomization procedure. As a result, the final QSAR equation after removal of outliers was as follows:

$$\log (PF) = 0.84 - 0.002 \times Hf + 0.01 \times Jurs_PNSA\text{-}1 - 0.01 \times Shadow_XZ$$

$$(N = 17;\ LOF = 0.015;\ r^2 = 0.962;\ r^2_{adj} = 0.953;\ F\text{-test} = 110.08;\ LSE = 0.006;\ r = 0.981;\ q^2 = 0.917)$$

where PF (potentiation factor) of the compounds (efflux pump inhibitors – EPI) express the reduction in the MIC values of ciprofloxacin, hence represent the potentiation of activity of ciprofloxacin in the presence of piperine analogs based efflux pump inhibitors; N is number of compounds in training set, LOF is lack of fit score, r^2 is squared correlation coefficient, r^2_{adj} is square of adjusted correlation coefficient, *F*-test is a variance-related static which compares two models differing by one or more variables to see if the more complex model is more reliable than the less com-

Figure 1. General structures of compounds taken for QSAR study



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

plex one, the model is supposed to be good if the *F*-test is above a threshold value, LSE is least-square error, *r* is correlation coefficient, q^2 is the square of the correlation coefficient of the cross-validation.

The QSAR model was proved robust by using the method proposed by Golbraikh *et al*., and Roy. The achieved QSAR model had the values of $R^2_{pred} = 0.78$ and $r^2_m = 0.72$, indicating the good external predictability of the QSAR model.

Jurs_PNSA-1, Shadow_XZ and Hf are important descriptors responsible for the relationship between piperine analogs' structure and activity as *S. aureus* NorA efflux pump inhibitors. Jurs_PNSA-1 is the partial negative surface area, calculated by using the sum of the solvent accessible surface area of all negatively charged atoms in the compound. Shadow_XZ is the area of the molecular shadow in the XZ plane. Hf is the heat of formation descriptor. These parameters were proved to be important contributors to the potentiating activity and must be taken into consideration while designing the efflux pump inhibitors.

QSAR of Aryl Alkenyl Amides/Imines

Another QSAR study on activity of aryl alkenyl amides/imines having structural similarity to piperine (structures shown in Figure 2) in inhibiting bacterial efflux pump was also conducted by Nargotra *et al* (A Nargotra et al., 2009).

The total set of 42 compounds was divided into 34 molecules in training set and 8 molecules in test set by arranging the molecules in ascending order of their PF and selecting every fifth molecule to be included in the training set. About 200 descriptors were calculated using the Cerius2 4.10 software package. After the removal of the descriptors with zero values for all the compounds, remaining descriptors were filtered using genetic function approximation (GFA) method. The study showed that the ideal number of descriptors should be five. After analyzing the equation with five descriptors and removing outliers, the following equation was achieved:

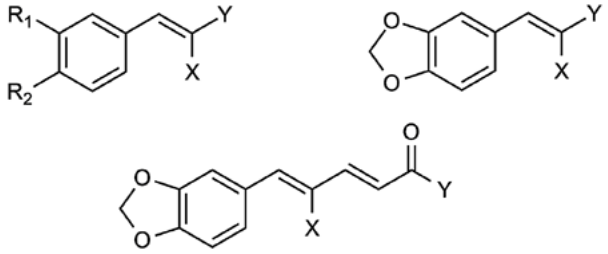
$$\log (PF) = -0.2028 - 0.000936*Hf + 0.342484*S_dssC$$

$$+ 0.441556*RadOfGyr - 0.2087*Atype\text{-}C\text{-}16$$

$$- 0.00104*Jurs\text{-}WNSA2$$

($N = 30$; LOF = 0.051; $r^2 = 0.866$; $r^2_{adj} = 0.833$; F-test = 28.875; LSE = 0.022; $r = 0.929$; $q^2 = 0.757$)

Figure 2. General structures of compounds taken for present study



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

The final QSAR model had the values of $R^2_{pred} = 0.825$ and $r^2_m = 0.634$, indicating the good external predictability of the QSAR model. The QSAR study also gives us the revelation that E-state indices and A log P atom type are important descriptors responsible for describing the activity of aryl alkenyl amide/imine based efflux pump inhibitors. In addition, solvent accessible surface area, partial charged surface area of a molecule, radius of gyration and heat of formation of the molecule are also important parameters to be taken into account while designing new inhibitors belonging to the class of compounds mentioned above.

QSAR of 2-Aryl-5-Nitro-1H-Indole Derivatives

It has been found that the combinatorial administration of lead structure 5-nitro-2-phenylindole (INF55) and the antibiotic significantly increases *S. aureus* susceptibility to ciprofloxacin (Ambrus et al., 2008). Structure-activity relationship studies mostly aimed at disclosing the importance of the C5 position of INF55 (Figure 3) for inhibitory activity against the NorA pump:

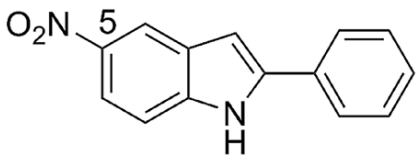
1. Reduction of C5 substituent will adversely affect the activity.
2. Carbonyl based electron-withdrawing groups at C5 eliminate all activity.
3. C5 substituents such as sulfonic acids/esters/amides have no effect on potentiation.
4. INF55 derivatives bearing a nitrile group show similar potentiation activity to INF55 (compound 14) (Ambrus et al., 2008).

All compounds lacking 5-NO_2 groups were significantly less potent than INF55. The 3′-CO_2Me INF55 derivative showed slightly higher potency than INF55 against all three *S.aureus* strains. INF55 analogue containing a 4′-CH_2OH group was nearly equipotent with INF55 against the wild type and NorA overexpressing strain, representing a promising candidate for incorporation into dual action mutual prodrugs targeting the NorA pump (Ambrus et al., 2008).

Berberine is the active constituent of the medicinal plants echinacea and golden seal. The extrusion of berberine by bacterial efflux pumps can make this antimicrobial agent ineffective. Nevertheless, it was found that the use of berberine along with an effective MDR pump inhibitor could restore its activity. In another related study, an inhibitor of Major Facilitator MDRs, namely SS14, was made on the basis of covalent link between berberine and INF55. Compared to berberine, SS14 was 100-fold more active against *S.aureus* and 200-400 fold more active against *S.aureus* mutants overexpressing the NorA MDR (Ball et al., 2006).

In a QSAR study on MDR inhibitory activities of indole derivatives, the 3D structures of ten 2-aryl-5-nitro-1-*H*-indole derivatives (structures shown in Table 2) were optimized with the semi-empirical PM3 method and density functional theory (DFT) at b3lyp/6-31 g* level successively (Dai, Zhang, Zhang, Wang, & Lu, 2008).

Figure 3. Structure of INF55



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

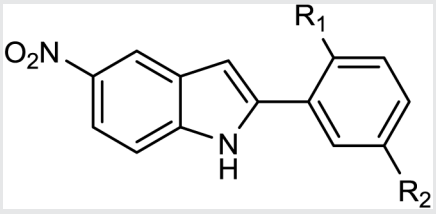
Some structural and electric parameters were subsequently obtained from single point energy and natural bond orbital analysis. These parameters included the lowest unoccupied molecular orbital energy (E_{LUMO}), the highest occupied molecular orbital energy (E_{HOMO}), the polarizability (α), the nuclear-nuclear repulsion energy (E_{rep}), the molecular total energy (E_{tot}), the dipole (μ), the molecular volume (V_{mol}), and the Mulliken atomic charges of the six carbons of the 2-aryl (Q_{C1}-Q_{C6}). A genetic algorithm-based multi-linear regression analysis (GA-MRA) was carried out to reduce the number of the structural and electric parameters originally from 13 parameters obtained from the quantum chemistry calculations, thus rendering the optimized set of descriptors for constructing the QSAR model. With each data set containing m variables, a QSAR model was built. Then the optimum combinations containing the same numbers of descriptors were screened out with GA method. For each m of 2-7, two combinations of the descriptors with the top square correlation coefficient R^2 are analyzed based on the trends of their R^2, R^2_{CV}, and F (a=0.05) values in accordance with changes of *m* values. The QSAR model was built based on partial least square (PLS) analysis.

The outcomes of genetic algorithm analysis indicated that the model containing five descriptors of μ, V_{mol}, Q_{C2}, Q_{C3} and Q_{C4} with the highest R^2_{CV} of 0.9941 and a good R^2 of 0.9982 could be considered as an ideal model for QSAR of 2-aryl-5-nitro-1-*H*-indole. These five descriptors were re-analyzed with partial least square (PLS) and the LOO cross validation. The QSAR model achieved with 5 independent variables is shown as follows:

$$-\lg \mathrm{MIC} = -0.397 \times \mu + 0.019 \times V_{mol} + 1.231 \times Q_{C2} + 9.976 \times Q_{C3} - 6.177 \times Q_{C4} - 0.198 \quad (*)$$

$$N = 10,\ R^2 = 0.998,\ R^2_{CV} = 0.996$$

Table 2. Structures and activities of ten 2-aryl-5-nitro-1H-indole derivatives used for QSAR analysis

Compound			MIC of NorA++ K2361 Plus Berberine (30µg/mL)
	R_1	R_2	
7a	COOH	H	50
8a	CH_2OH	H	12.5
9a	CH_2N_3	H	3
7b	COOH	OCH_3	50
8b	CH_2OH	OCH_3	6.25
9b	CH_2N_3	OCH_3	1.5
8c	CH_2OH	OCH_2Ph	0.8
10a	CH_2NH_2	H	12.5
10b	CH_2NH_2	OCH_3	12.5
INF55	H	H	3

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

The equation showed that increasing Q_{C3} and decreasing Q_{C4} could improve the activity of 2-aryl-5-nitro-1-*H*-indole derivatives as efflux pump inhibitor. With the model achieved, the electric and structural descriptors of a new compound, 2-(2-azidomethyl-5-phenoxy-phenyl)-5-nitro-1*H*-indole (8d, R_1 = -CH_2N_3, R_2 = -OCH_2Ph) were calculated with the same quantum chemistry calculation method and they were predicted from the equation (*) to lower the MIC of berberine to 0.091 μg/ml for inhibiting K2361 of *S.aureus* with NorA efflux pump protein overexpression.

Case Study of Using QSAR Model for Virtual Screening of NorA Efflux Pump Inhibitors

In a QSAR study conducted by Thai *et al.* (Thai et al., 2015), a total of 47 compounds belonging to different structural classes in company with the efflux pump inhibitory (EPI) activity against NorA mediated ethidium bromide efflux (IC_{50}) were compiled from the literature (German et al., 2008; Kaatz, Moudgal, Seo, Hansen, et al., 2003; Kaatz, Moudgal, Seo, & Kristiansen, 2003; Michalet et al., 2007; Pieroni et al., 2010; Sabatini et al., 2012; Sabatini, Kaatz, Rossolini, Brandini, & Fravolini, 2008; Smith, Williamson, Wareham, Kaatz, & Gibbons, 2007; Stavri, Piddock, & Gibbons, 2007). In MOE, all chemical structures of the data sets were built and energy minimized before computing 184 2D molecular descriptors. Subsequently, the variables were selected using RapidMiner (RapidMiner), Weka (WEKA) and MOE (MOE) software packages respectively. This attribute selection resulted in five descriptors which were analyzed in correlation matrix with experimental pIC_{50} values. Subsequently, linear regression models ($y=a_0+a_ix_i$) were generated by using the QuaSAR-Model tool in MOE (MOE) with the method Partial Least Square (PLS) and the activity field pIC_{50}. Based on the training set, different models generated from different descriptors were internally validated by the cross-validation technique and externally validated by the test set. The descriptors used for the best one were continuously used for modeling the whole data set to get the final model which was internally validated by the cross-validation technique and externally validated by the external test set. As a result, the finally achieved model with statistical values $Q^2 = 0.8$, $\overline{r_m^2} = 0.71$ and $\Delta r_m^2 = 0.18$ met validation criteria for the data set following Roy *et al.* (Ojha, Mitra, Das, & Roy, 2011; Pratim Roy, Paul, Mitra, & Roy, 2009; Roy et al., 2011).

The yielded model was applied on drug-likeness compounds from the mixture library of 182 flavonoids (Belsare, Pal, Kazi, Kankate, & Vanjari, 2010; Farkas, Jakus, & Héberger, 2004; Park et al., 2006; Phosrithong, Samee, Nunthanavanit, & Ungwitayatorn, 2012; Seyoum, Asres, & El-Fiky, 2006; Tran, Park, Kim, Ecker, & Thai, 2009; Yokozawa et al., 1998) and the traditional Chinese medicine (C. Y. Chen, 2011) (TCM) database that met the rules of five/three to identify novel Nor-A inhibitors with $pIC_{50} \geq 4.5$. Among 182 flavonoids having the rings values ≥ 2, 27 of 169 compounds that matched the criteria of the rule of five and none of 8 compounds that matched the criteria of the rule of three were predicted as potential NorA inhibitors. Among 52957 TCM compounds with the descriptor rings ≥ 2, 2866 of 15250 compounds that matched the criteria of the rule of five and 9 of 1679 compounds that matched the criteria of the rule of three were determined as hit compounds for NorA inhibitory activity as well. Ultimately, 27 flavonoids and 2863 TCM compounds had fully satisfied at least the Lipinski's rule of five and had good predicted pIC_{50} values and thus were considered as potential drugs to reverse the efflux-mediated multidrug resistance of *Staphylococcus aureus*.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Additional Discussion

From QSAR studies on piperine, aryl alkenyl amides/imines and 2-aryl-5-nitro-1*H*-indole derivatives as ligands of NorA efflux pump, it was observed that certain portion of any kind of ligands do involve themselves in hydrophobic interaction or electronic or in both with the receptor. The crucial binding in most of the cases is, nonetheless, hydrogen bonds. NorA is, therefore, assumed to have certain critical features capable of interacting with all kinds of structurally different ligands. These features may comprise some hydrogen-bond donor and acceptor sites, some lipophilic areas, and a few polar sites.

This view was justified by two docking studies using NorA homology models: Thai *et al*. developed homology model of NorA based on the transporter (PDB ID: 2gfp) while Kalia *et al*. used GlpT (PDB ID: 1PW4) as the protein template to create homology model. Thai *et al*. docked 6 traditional Chinese medicine (TCM) and 27 flavonoids into the central channel and the Walker B. Most molecules had better affinities to the Walker B than the central channel due to the hydrophilic property of the latter. Compared with TCM, flavonoids have higher affinities towards NorA binding site because of tight interactions (hydrogen bonding and π-π stacking) with NorA homology model made by substituents in flavonoid structures (Thai et al., 2014). Kalia *et al*. conducted docking studies of two ligands namely reserpine and capsaicin, which are two known NorA inhibitors, in the binding site of NorA. The most important finding was that a hydrophobic cleft of NorA provided stability as well as affected the orientation of the ligand within the binding pocket. Hydrogen bondings between substituents of ligands and amino acid residues of the binding site attributed to the stability of the ligand-protein complex (Kalia et al., 2012). These results consolidated the outcomes of above-presented QSAR studies that most of parameters selected for final models are associated with surface charge, polarity and molecular weight.

Compared with QSAR studies of specific derivatives, the QSAR analysis of natural products for NorA EPIs was performed with a larger and more diversified database including the compounds from different structural classes. However, this strong point became an obstacle that made the validation results of obtained models be at a disadvantage compared with those of the model generated by Koul *et al.* (A. Nargotra et al., 2009). Among three descriptors used for the model of these authors (A. Nargotra et al., 2009), this study revealed that the descriptor "partial negative surface area" might be relevant to the descriptor "total negative vdw surface area" (PEOE_VSA_NEG) that was determined as a less important variable and removed from final models.

QSAR Approaches in Search for EPIs in Gram Negative Bacteria

Gram-negative bacteria seem to be more dangerous than Gram-positive bacteria in some cases. Some can be named such as *P. aeruginosa, K. pneumoniae*, *C. jejuni*, *E. coli*, *S. typhimurium* and *E. aerogenes, etc.* Moreover, their levels of MDR have increased alarmingly. Nonetheless, no efflux pump inhibitor has been licensed for use in the treatment of bacterial infections in human or veterinary medicine. It must stimulate research that leads to the development of new EPI molecules.

With the recent advent of high-resolution 3D-structure of drug transporters, the co-crystallization of AcrB and MexB with some substrates, and the analyses of the biological effects of several mutations in the pump, it is now possible to rationally design functional targets located inside the bacterial pumps. Consequently, the synthesis of efficient molecules with strong EPI capacity will be more readily achieved. In this context, there should be more QSAR models that help to save time and money on searching "new

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

drugs" for mediation of efflux pumps in Gram-negative bacteria. However, it is noticed with deep concern that scientists have not paid enough attention on this kind of QSAR studies. In comparison with a large number of other efflux pumps in Gram-negative bacteria, studies on the two kinds of efflux pumps (MATE efflux pump VmrA from *V.parahaemolyticus*, and AcrAB–TolC from *Escherichia coli*) are likely insufficient to deal with Gram-negative bacterial resistance. Therefore, it is hoped that new and solid QSAR models for MDR efflux proteins would be yielded in order to provide a more complete picture of the molecular requirements.

QSAR Study of β-Lactam Antibiotic Efflux by the Bacterial MDR Pump AcrB

A QSAR study related to Gram-negative bacterial efflux pumps was conducted by researchers Marcia Ferreira and R.Kiralj in 2004 (Ferreira & Kiralj, 2004). The target of the study was AcrAB-TolC – an efflux pump identified in several Gram-negative bacteria including *Salmonella typhymurium*. AcrAB-TolC is a tripartite efflux pump, which consists of the inner-membrane transporter AcrB (Edward, McDermott, Zgurskaya, Nikaido, & Koshland, 2003; Elkins & Nikaido, 2003; Murakami, Nakashima, Yamashita, & Yamaguchi, 2002; Pos & Diederichs, 2002), the outer-membrane channeltunnel protein (Koronakis et al., 2000) of the TolC family (Elkins & Nikaido, 2003), and the periplasmic linker lipoprotein AcrA from the membrane fusion family (Thanabalu, Koronakis, Hughes, & Koronakis, 1998).

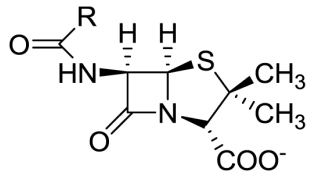
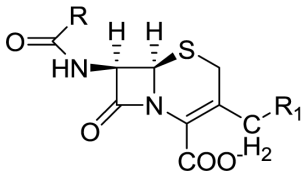
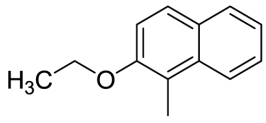
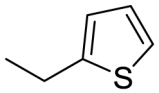
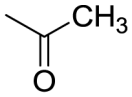
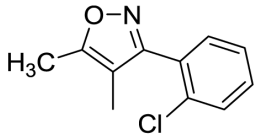
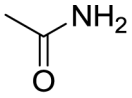
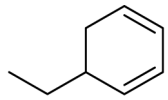
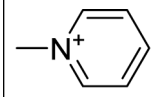
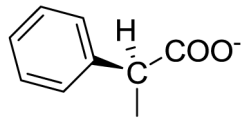
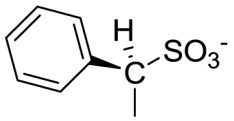
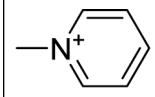
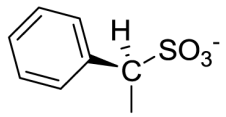
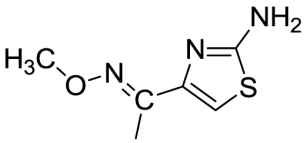
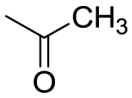
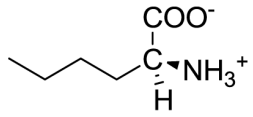
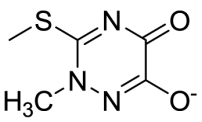
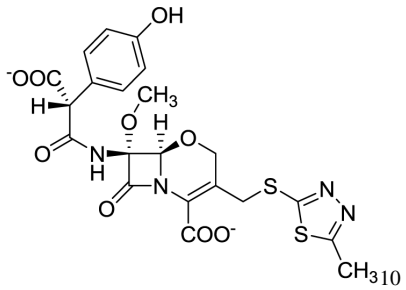
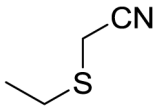
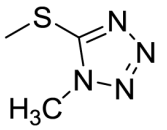
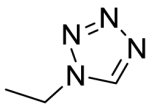
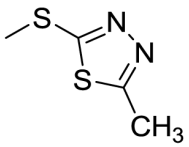
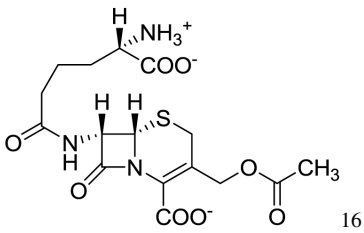
The study was aimed at establishing multivariate quantitative lipophilicity–MIC and β-lactam structure–MIC relationships for *S. typhimurium* strains (Nikaido, Basina, Nguyen, & Rosenberg, 1998). Mass concentration MICs for 16 penicillines and cephalosporins (structures shown in Table 3) with respect to bacterial strains *S. typhimurium* SH5014 (parent strain) and its mutants low resistant SH7616 and AcrAB overproducer HN891, were collected from the literature (Nikaido et al., 1998). Around 50 molecular properties of 16 β-lactam antibiotics were calculated using different methods based on their modeled molecular structures. These computed properties included lipophilicity parameters and other molecular properties which were prepared based on two-dimensional chemical formulae of studied penicillins and cephalosporins. HCA (hierarchical cluster analysis) and PCA (principal component analysis) studies indicated that the presence of charged groups and hydrophobic moieties determined the antibiotic behavior with respect to AcrAB-TolC pumps. Highly charged β-lactams with small hydrophobic fraction w_C or S_f might be poor AcrB substrates due to energetically unfavorable pump–drug interaction in molecular recognition. Nonetheless, the HCA-PCA approaches could not give answers to the question of which lipophilicity parameters were most reliable for these antibiotics, so the QSAR studies were carried out.

QSAR studies, which were performed by means of the partical least squares (PLS) and based on lipophilicity and electronic and hydrogen bonding molecular descriptors, demonstrated that

1. Biological activities (pMICs) strongly depend both on properties of bacterial strains and drug molecules. β-lactams were classified as good, moderately good and poor AcrAB-TolC substrates;
2. Among the most important β-lactam molecular properties quantitatively related to pMICs are lipophilicity and electronic and hydrogen bonding properties; and
3. Lipophilicity parameters calculated in different ways do not necessarily present the same information about drugs, and cannot produce parsimonious and accurate regression models for MICs originated by active AcrAB-TolC pumps.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 3. Chemical structures of β-lactams used in QSAR study

No.	R	No.	R	R_1
				
1		4		
2		5		
3		6		
7		9		
8		11		
15		12		
 10		13		
		14		
		 16		

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Two strains HN981 and SH5014 were found to act similarly, and the third one – SH7616 is significantly different from each of them (Table 4). This was manifested by the similarity, even with the same molecular descriptors, of the PLS models related to strains HN891 and SH5014.

Additionally, penicillins and cephalosporins which can be characterized as hydrophilic, with good hydrogen bonding properties and able to establish polar-polar interactions with bacterial pump receptors, are bad pump substrates, and so potentially good drugs. This finding has been supported by recent studies of AcrB in which most pharmacophore models generated by 49 AcrB substrates contain hydrophobic centers and hydrogen bond acceptors or donors. Molecular interactions of phytochemicals with the active site residues in the binding site of AcrB also consist of hydrogen bondings and hydrophobic contacts (Aparna, Dineshkumar, et al., 2014). Vargiu and Nikaido confirmed the similar features of AcrB binding sites. Distal pocket and proximal pocket were shown to be two main active regions of AcrB. 15 hydrophobic residues and 11 polar or charged amino acid residues in the distal binding pocket were found to contribute to the binding of all ligands examined (Vargiu & Nikaido, 2012). In another study publicized on *Nature* magazine in 2013, scientists found that AcrAB-MexAB specific inhibitor of pyridopyrimidine derivative (ABI-PP) is deeply inserted into a hydrophobic trap rich in phenylalanine residues. Based on the results of site-directed mutagenesis and molecular dynamics simulations, the trap has been considered to indirectly affect drug-binding site. Phe178 located at the edge of this trap in AcrB contributes to the tight binding of the inhibitor molecule through a π-π interaction with the pyridopyrimidine ring (Nakashima et al., 2013).

QSAR/SAR Study of Structurally Unrelated Substrates of a MATE Efflux Pump VmrA from V.Parahaemolyticus

Multidrug transporters of MATE family have been known to expel dissimilar lipophilic and cationic drugs across cell membranes by dissipating a preexisting Na+ or H+ gradient (Lu et al., 2013). Brown *et al*. classified the MATE family into three subfamilies. Subfamily 1 includes NorM, subfamily 2 includes ERC1, and subfamily 3 includes DinF. VmrA has been considered to be a member of subfamily 3 (DinF branch). The *vmrA* gene consisted of 1,341 nucleotides, with a deduced popypeptide (VmrA) consisting of 447 amino acids residues with a calculated molecular mass of 49 kDa. VmrA was very rich in hy-

Table 4. PLS regression models for pMICs

pMIC	Parameters[a]	SEP[b]	Q[c]	R[d]	PCs (%)
HN891	w_C, S_f, GlogK_{OW}, logP_s, SlogP_s, logK_{WIN}, SlogK_{WIN}, logP_{IA}, logP_X	0.467	0.912	0.967	3 (82)
	logP_{IA}, GlogK_{OW}, SlogP_s, H_f, A_{HB}, D_y, N_{NS}	0.209	0.982	0.993	3 (85)
SH5014	w_C, S_f, GlogK_{OW}, logP_s, SlogP_s, logK_{WIN}, SlogK_{WIN}, logP_{IA}, logP_X	0.391	0.942	0.975	2 (82)
	logP_{IA}, GlogK_{OW}, SlogP_s, H_f, A_{HB}, D_y, N_{NS}	0.316	0.962	0.982	3 (85)
SH7616	w_C, S_f, GlogK_{OW}, logP_s, SlogP_s, logK_{WIN}, SlogK_{WIN}, logP_{IA}, logP_X	0.792	0.645	0.886	2 (76)
	logP_{IA}, GlogK_{OW}, SlogK_{WIN}, N_{CH}, H_f, N_{NS}	0.461	0.851	0.930	3 (83)

[a]Transformation of lipophilicity descriptors;[b] Standard error of validation; [c] Correlation coefficient from validation; [d] Correlation coefficient from prediction.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

drophobic residues, indicating that the protein was an integral membrane protein (J. Chen et al., 2002). This characterization of MATE pump favoring hydrophobic substrates were illustrated in a QSAR study of VmrA substrates done by Rudolf Kiralj and Marcia M.C. Ferreira in 2008 (Kiralj & Ferreira, 2008).

Kiralj and Ferreira aimed at correlating the efflux activity of *V. parahaemolyticus* VmrA with molecular properties of 12 diverse agents at QSAR and SAR levels. The training set in this work consists of 12 agents (organic molecules/ions including 4',6-diamino-2-2phenylindole, tetraphenylphosphonium chloride, acriflavine chloride, ethidium bromide, chloramphenicol, norfloxacin, rhodamine 6G chloride, tetracycline, erythromycin, streptomycin, sodium deoxycholate and sodium dodecyl sulfate). The corresponding activities pMIC = -log(MIC/mol/dm^3) are pMIC(KAM) and pMIC(pVCJ6), which are efflux activity pMIC of the *E.coli* strains KAM32 and KAM32/pVCJ6, respectively. The prediction of the MDR character of VmrA with respect to 19 diverse substrates was performed at both QSAR and SAR levels.

Analysis of descriptors in correlation with the biological activities pMIC indicates that MDR of VmrA was directed against agents with shorter bonds, a few polar atoms, elevated contents of rings, planar structures, and aromatic carbon atoms. This is generally a rigid aromatic structure with exo- and/or endocyclic heteroatoms. Satisfactory prediction power statistics $Q^2>0.5$ and $R^2>0.6$ for QSAR models can be observed for all the models. The low R^2 and Q^2 values indicated that the good results in the original models were not due to a chance correlation or structural dependency of the training set.

There were also common structural features related to efflux from KAM32 and KAM32/pVCJ6 strains. In general, agent efflux was not determined by specific functional groups due to the strictly nonbonding nature of pump-mediated MDR. Exploratory analysis related to the two strains points out that good substrates of efflux systems in *E.coli* strains KAM32 and KAM32/pVCJ6 are neither linear nor extended, preferably were ring structures with some branching, rather hydrophobic, weak dipoles, difficult to polarize, possess some polar groups, and could establish hydrogen bonds with receptors. The strains differed in their preferences for good substrates: KAM32 binds rather hydrophobic species, while KAM32/pVCJ6 attracts heteroaromatics. Therefore, aromatic-aromatic and hydrophobic-hydrophobic VmrA-substrate interactions seemed to be crucial. Agent-receptor interaction statistics show that agent-receptor interaction descriptors depend on agent molecular properties, receptor molecular properties, agent position and binding mode relative to the receptor, and the presence of other molecular/ionic species. Supposing that agent properties might substantially affect agent-receptor interactions, variations in receptors and agent binding modes/positions were expected to have a secondary importance. In addition, good substrates of VmrA efflux pump from *V. parahaemolyticus* strain AQ3334 must also be rather rigid and condensed heteroaromatic systems, with very few or no flexible side-chains, with exo- or endocyclic heteroatoms, and establish aromatic-aromatic contacts with VmrA.

This QSAR study proved that chemometric methods are very useful in practice and the agent-receptor interaction descriptors, based on PDB data, correlate reasonably with biological activities, further rationalizing the VmrA-agent interactions.

DESCRIPTORS AND METHODS

It has been recognized that most descriptors employed to build QSAR models for prediction of interactions between chemical agents and Gram-negative bacterial efflux pumps were mostly parameters of molecular field (steric, electrostatic and hydrophobic). In the study of AcrAB-TolC, critical β-lactam molecular descriptors related to pMICs were associated with lipophilicity. For the examination of VmrA

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

substrates, VmrA in strain KAM32 extrudes agents characterized by three lipophilicity descriptors (Nh, sighyd and wh2) and two electronic descriptors (E4 and HOMO) while the regression vectors for KAM32/pVCJ6 have contributions of three electronic descriptors (FF, Np2 and Mrefn2), a steric (wr) and a hydrophobic (sighyd) descriptors accounting for agent efflux by VmrA. In studies of MDR efflux pumps of Gram-positive bacteria, in addition to molecular field descriptors as main contributors to QSAR models, other relevant descriptors relate to quantum chemical (heat of formation), constitutional (number of rings), topological and geometrical (molecular volume and radius of gyration) properties. These observations were consistent with the fact that even though the mechanisms of drug transport in bacterial efflux pumps have remained unknown and controversial, empiric observations on properties of known pump substrates and inhibitors have indicated several regular features, of which amphiphilic nature of compounds play a key role. Further discussion regarding this character of MDR pump inhibitors can be found in the review and the commentary done by Van Bambeke *et al*. in which the presence of well-defined lipophilic and hydrophilic regions of antibiotics was mentioned (Van Bambeke, Balzi, & Tulkens, 2000; Van Bambeke & Lee, 2006).

All QSAR studies concerning NorA inhibitors included in this chapter principally relied on two-dimensional (2D) descriptors to generate the models. Among the used methods, the most common was PLS, others included PCR and HCA. However, 2D descriptors could not give relevant information regarding the ligand – efflux pump protein interaction. Besides, there was only a limited number of studied transport proteins, as the researchers focused on several efflux pumps, namely NorA, AcrAB-TolC, and VmrA, out of a large number of identified antibiotic-resistant efflux pumps mentioned in previous sections of this chapter. Therefore, the need of broadening the scale of QSAR research and studies based on bacterial efflux pumps is evident.

FUTURE RESEARCH AND OUTLOOK

As far as we are concerned, since the parameters used for presented 2D-QSAR models are mostly molecular field descriptors, the COMFA/COMSIA-based 3D-QSAR might be an appropriate alternative for the exploration of relationships between certain structural properties and the binding mode in which the ligand interacts with the efflux pump at the active domains. However, this method also has several limitations. The alignment protocols used in 3D-QSAR assumes that the ligands observed in the model bind to the efflux pump at the same binding site in the same manner. As a result, researchers may have difficulties achieving 3D-QSAR models in the circumstances where the efflux pumps have either one large binding site or multiple binding sites which can accommodate simultaneous ligands in various ways. In such cases, VolSurf/GRIND-based descriptors should be employed to obtain 3D-QSAR models with acceptable predictive capability. These are alignment free and thus allow the analysis of structurally diverse compound sets. Although QSAR approaches do not enable one to decipher the underlying mechanism of ligand-protein interaction, they might represent versatile tools for *in silico* filtering large combinatorial libraries for compounds with bacterial efflux pump inhibitory activity.

Moreover, the studies mentioned in the chapter mainly relied on the PLS analysis to yield predictive models. This approach was limited to sets of compounds whose biological activity and molecular descriptors are in linear correlations. Thus, there should be QSAR studies using machine learning and neuron

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

network methods which can deal with non-linear structure-activity relationships. These approaches may pave the way for the establishment of rapid *in silico* screening that are routinely applied in the early drug discovery phase. Additionally, pharmacophore modeling should also be considered in such the cases where the database includes a large number of compounds that have diverse structural properties. This method may be used alongside 3D-QSAR in order to assess the drug – transporter interactions, and also to give certain suggestions on the direction of synthesizing new chemical compounds to inhibit the studied efflux pumps. The pharmacophore models can be derived from either the binding site of the protein or a group of active compounds (Guner, 2000; Mannhold, Kubinyi, Folkers, Langer, & Hoffmann, 2006).

Recently, many researchers are working on the efflux pump namely NorA of *S. aureus*. However, no X-ray crystal structures have been found until now. Hence, physiological, pharmacological and medicinal chemistry studies on NorA pump have contributed to our understanding of the structure and mechanism of drug resistance. Further *in silico* as well as *in vitro* studies and X-ray crystal structures of certain efflux pumps like NorA have remained attractive targets to scientist working in intelligent designs of new novel NorA inhibitors as well as other EPIs.

CONCLUDING REMARKS

This review provides a summary of current knowledge about efflux-pump-mediated multidrug resistance and QSAR studies on bacterial EPIs. QSAR modeling results provide useful information for structural modifications of existing compounds in order to gain more potent compounds in use. The major benefits derived from developing efficient EPIs will be the ability to reuse various antibiotics affected by the efflux pumps as well as the control of the emergence and the dissemination of MDR efflux strains. Until now, the majority of the known inhibitors have been obtained from screening libraries of synthetic compounds, from purification of natural compounds or chemical modification of existing molecules.

As far as Gram-positive bacteria are concerned, most of the QSAR studies on bacterial MDR are dedicated to NorA in *S.aureus*. Obtained QSAR analyses, coupled with molecular docking studies, proved that hydrophobicity and hydrogen bondings are essential for NorA inhibitory activity. For Gram-negative bacterial EPIs, there has been a modest number of QSAR studies conducted. With the advent of high-resolution crystal structures of efflux pumps of *Escherichia coli* and *Pseudomonas aeruginosa*, it is hoped that consequent increase in the population of identified molecules with high EPI capacity would foster ligand-based approaches in this area. For general QSAR studies on bacterial EPIs, 2D-QSAR using partial least squares method and molecular field parameters were largely employed. Therefore, it is suggested that 3D-QSAR and machine learning methods should be taken into consideration when generating future QSAR models. Also, additional X-ray structures of proteins will definitely aid in the understanding of the molecular principles underlying the ligand-protein binding mode of bacterial efflux pumps.

To date, none of bacterial EPIs have entered clinical trials yet, the active compounds found with EPI-potency should still be subjects of deeper examination, including their pharmacokinetic properties and toxicity, as they give a hope for their future therapeutic usage. Moreover, the development of combination therapy poses additional complexity due to the necessity of precision tailoring of the pharmacokinetics of both agents to achieve the desired pharmacodynamic effect.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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).

REFERENCES

Akama, H., Kanemaki, M., Yoshimura, M., Tsukihara, T., Kashiwagi, T., & Yoneyama, H. et al. (2004). Crystal structure of the drug discharge outer membrane protein, OprM, of *Pseudomonas aeruginosa* dual modes of membrane anchoring and occluded cavity end. *The Journal of Biological Chemistry*, *279*(51), 52816–52819. doi:10.1074/jbc.C400445200 PMID:15507433

Alekshun, M., & Nelson, M. (2004). *WO patent no. 2004062674*.

Ambrus, J. I., Kelso, M. J., Bremner, J. B., Ball, A. R., Casadei, G., & Lewis, K. (2008). Structure–activity relationships of 2-aryl-1*H*-indole inhibitors of the NorA efflux pump in *Staphylococcus aureus*. *Bioorganic & Medicinal Chemistry Letters*, *18*(15), 4294–4297. doi:10.1016/j.bmcl.2008.06.093 PMID:18632270

Aparna, V., Dineshkumar, K., Mohanalakshmi, N., Velmurugan, D., & Hopper, W. (2014). Identification of natural compound inhibitors for multidrug efflux pumps of *Escherichia coli* and *Pseudomonas aeruginosa* using *In Silico* high-throughput virtual screening and *In Vitro* validation. *PLoS ONE*, *9*(7), e101840. doi:10.1371/journal.pone.0101840 PMID:25025665

Aparna, V., Mohanalakshmi, N., Dineshkumar, K., & Hopper, W. (2014). Identification of inhibitors for RND efflux pump of *Pseudomonas aeruginosa* using structure-based pharmacophore modeling approach. *International Journal of Pharmacy & Pharmaceutical Sciences*, *6*(1).

Atal, C., Zutshi, U., & Rao, P. (1981). Scientific evidence on the role of Ayurvedic herbals on bioavailability of drugs. *Journal of Ethnopharmacology*, *4*(2), 229–232. doi:10.1016/0378-8741(81)90037-4 PMID:7311598

Bachas, S., Eginton, C., Gunio, D., & Wade, H. (2011). Structural contributions to multidrug recognition in the multidrug resistance (MDR) gene regulator, BmrR. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(27), 11046–11051. doi:10.1073/pnas.1104850108 PMID:21690368

Ball, A. R., Casadei, G., Samosorn, S., Bremner, J. B., Ausubel, F. M., Moy, T. I., & Lewis, K. (2006). Conjugating berberine to a multidrug resistance pump inhibitor creates an effective antimicrobial. *ACS Chemical Biology*, *1*(9), 594–600. doi:10.1021/cb600238x PMID:17168555

Bavro, V. N., Pietras, Z., Furnham, N., Pérez-Cano, L., Fernández-Recio, J., & Pei, X. Y. et al. (2008). Assembly and channel opening in a bacterial drug efflux machine. *Molecular Cell*, *30*(1), 114–121. doi:10.1016/j.molcel.2008.02.015 PMID:18406332

Bay, D. C., Rommens, K. L., & Turner, R. J. (2008). Small multidrug resistance proteins: A multidrug transporter family that continues to grow. *Biomembranes*, *1778*(9), 1814–1838. doi:10.1016/j.bbamem.2007.08.015

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Belsare, D. P., Pal, S. C., Kazi, A. A., Kankate, R. S., & Vanjari, S. S. (2010). Evaluation of antioxidant activity of chalcones and flavonoids. *International Journal of ChemTech Research*, *2*(2), 1080–1089.

Bhardwaj, R. K., Glaeser, H., Becquemont, L., Klotz, U., Gupta, S. K., & Fromm, M. F. (2002). Piperine, a major constituent of black pepper, inhibits human P-glycoprotein and CYP3A4. *The Journal of Pharmacology and Experimental Therapeutics*, *302*(2), 645–650. doi:10.1124/jpet.102.034728 PMID:12130727

Borges-Walmsley, M., McKeegan, K., & Walmsley, A. (2003). Structure and function of efflux pumps that confer resistance to drugs. *The Biochemical Journal*, *376*(2), 313–338. doi:10.1042/BJ20020957 PMID:13678421

Bush, K., & Miller, G. H. (1998). Bacterial enzymatic resistance: β-lactamases and aminoglycoside-modifying enzymes. *Current Opinion in Microbiology*, *1*(5), 509–515. doi:10.1016/S1369-5274(98)80082-9 PMID:10066532

Chamberland, S., Ishida, Y., Lee, V. J., Leger, R., Nakayama, K., Ohta, T., et al. (2002). *US patent no. 6399629*. Washington, DC: US Patent Office.

Chang, C., Ray, A., & Swaan, P. (2005). *In silico* strategies for modeling membrane transporter function. *Drug Discovery Today*, *10*(9), 663–671. doi:10.1016/S1359-6446(05)03429-X PMID:15894231

Chen, C. Y. (2011). TCM Database@Taiwan: The world's largest traditional Chinese medicine database for drug screening in silico. *PLoS ONE*, *6*(1), e15939. doi:10.1371/journal.pone.0015939 PMID:21253603

Chen, J., Morita, Y., Huda, M. N., Kuroda, T., Mizushima, T., & Tsuchiya, T. (2002). VmrA, a member of a novel class of Na^+-coupled multidrug efflux pumps from *Vibrio parahaemolyticus. Journal of Bacteriology*, *184*(2), 572–576. doi:10.1128/JB.184.2.572-576.2002 PMID:11751837

Chevalier, J., Bredin, J., Mahamoud, A., Malléa, M., Barbe, J., & Pagès, J.-M. (2004). Inhibitors of antibiotic efflux in resistant Enterobacter aerogenes and Klebsiella pneumoniae strains. *Antimicrobial Agents and Chemotherapy*, *48*(3), 1043–1046. doi:10.1128/AAC.48.3.1043-1046.2004 PMID:14982806

Chung, Y., & Saier, M. Jr. (2001). SMR-type multidrug resistance pumps. *Current Opinion in Drug Discovery & Development*, *4*(2), 237–245. PMID:11378963

Coban, A. Y., Ekinci, B., & Durupinar, B. (2004). A multidrug efflux pump inhibitor reduces fluoroquinolone resistance in *Pseudomonas aeruginosa* isolates. *Chemotherapy*, *50*(1), 22–26. doi:10.1159/000077280 PMID:15084801

Dai, Y., Zhang, X., Zhang, X., Wang, H., & Lu, Z. (2008). DFT and GA studies on the QSAR of 2-aryl-5-nitro-1H-indole derivatives as NorA efflux pump inhibitors. *Journal of Molecular Modeling*, *14*(9), 807–812. doi:10.1007/s00894-008-0328-6 PMID:18575902

Delmar, J. A., Su, C.-C., & Yu, E. W. (2014). Bacterial multidrug efflux transporters. *Annual Review of Biophysics, 43*, 93-117. doi: 10.1146/annurev-biophys-051013-022855

DePristo, M. A., de Bakker, P. I., & Blundell, T. L. (2004). Heterogeneity and inaccuracy in protein structures solved by X-ray crystallography. *Structure (London, England)*, *12*(5), 831–838. doi:10.1016/j.str.2004.02.031 PMID:15130475

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Du, D., Wang, Z., James, N. R., Voss, J. E., Klimont, E., & Ohene-Agyei, T. et al. (2014). Structure of the AcrAB-TolC multidrug efflux pump. *Nature*, *509*(7501), 512–515. doi:10.1038/nature13205 PMID:24747401

Edward, W. Y., Aires, J. R., McDermott, G., & Nikaido, H. (2005). A periplasmic drug-binding site of the AcrB multidrug efflux pump: A crystallographic and site-directed mutagenesis study. *Journal of Bacteriology*, *187*(19), 6804–6815. doi:10.1128/JB.187.19.6804-6815.2005 PMID:16166543

Edward, W. Y., McDermott, G., Zgurskaya, H. I., Nikaido, H., & Koshland, D. E. (2003). Structural basis of multiple drug-binding capacity of the AcrB multidrug efflux pump. *Science*, *300*(5621), 976–980. doi:10.1126/science.1083137 PMID:12738864

Elkins, C. A., & Nikaido, H. (2003). 3D structure of AcrB: The archetypal multidrug efflux transporter of *Escherichia coli* likely captures substrates from periplasm. *Drug Resistance Updates*, *6*(1), 9–13. doi:10.1016/S1368-7646(03)00004-9 PMID:12654283

Farkas, O., Jakus, J., & Héberger, K. (2004). Quantitative structure–antioxidant activity relationships of flavonoid compounds. *Molecules (Basel, Switzerland)*, *9*(12), 1079–1088. doi:10.3390/91201079 PMID:18007505

Ferreira, M., & Kiralj, R. (2004). QSAR study of β-lactam antibiotic efflux by the bacterial multidrug resistance pump AcrB. *Journal of Chemometrics*, *18*(5), 242–252. doi:10.1002/cem.867

German, N., Kaatz, G. W., & Kerns, R. J. (2008). Synthesis and evaluation of PSSRI-based inhibitors of *Staphylococcus aureus* multidrug efflux pumps. *Bioorganic & Medicinal Chemistry Letters*, *18*(4), 1368–1373. doi:10.1016/j.bmcl.2008.01.014 PMID:18242086

Guner, O. F. (2000). *Pharmacophore perception, development, and use in drug design*. La Jolla, CA: International University Line.

Handzlik, J., Matys, A., & Kieć-Kononowicz, K. (2013). Recent advances in multi-drug resistance (MDR) efflux pump inhibitors of Gram-positive bacteria *S. aureus*. *Antibiotics*, *2*(1), 28–45. doi:10.3390/antibiotics2010028

He, X., Szewczyk, P., Karyakin, A., Evin, M., Hong, W.-X., Zhang, Q., & Chang, G. (2010). Structure of a cation-bound multidrug and toxic compound extrusion transporter. *Nature*, *467*(7318), 991–994. doi:10.1038/nature09408 PMID:20861838

Higgins, C. F. (2001). ABC transporters: Physiology, structure and mechanism–an overview. *Research in Microbiology*, *152*(3), 205–210. doi:10.1016/S0923-2508(01)01193-7 PMID:11421269

Hinchliffe, P., Greene, N. P., Paterson, N. G., Crow, A., Hughes, C., & Koronakis, V. (2014). Structure of the periplasmic adaptor protein from a major facilitator superfamily (MFS) multidrug efflux pump. *FEBS Letters*, *588*(17), 3147–3153. doi:10.1016/j.febslet.2014.06.055 PMID:24996185

Hollenstein, K., Dawson, R. J., & Locher, K. P. (2007). Structure and mechanism of ABC transporter proteins. *Current Opinion in Structural Biology*, *17*(4), 412–418. doi:10.1016/j.sbi.2007.07.003 PMID:17723295

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Hvorup, R. N., Winnen, B., Chang, A. B., Jiang, Y., Zhou, X. F., & Saier, M. H. Jr. (2003). The multidrug/oligosaccharidyl-lipid/polysaccharide (MOP) exporter superfamily. *European Journal of Biochemistry*, *270*(5), 799–813. doi:10.1046/j.1432-1033.2003.03418.x PMID:12603313

Inoko, H., & Yoshihara, E. (2011). *7985410*. Washington, DC: US Patent Office.

Jack, D. L., Yang, N. M., & Saier, H, M. (. (2001). The drug/metabolite transporter superfamily. *European Journal of Biochemistry*, *268*(13), 3620–3639. doi:10.1046/j.1432-1327.2001.02265.x PMID:11432728

Jarmuła, A., Obłąk, E., Wawrzycka, D., & Gutowicz, J. (2010). Oporność wielolekowa związana z aktywnym usuwaniem leków z komórek drobnoustrojów. *Postepy higieny i medycyny doswiadczalnej (Online)*, *65*, 216–227. doi:10.5604/17322693.937011 PMID:21502698

Kaatz, G. W., Moudgal, V. V., Seo, S. M., Hansen, J. B., & Kristiansen, J. E. (2003). Phenylpiperidine selective serotonin reuptake inhibitors interfere with multidrug efflux pump activity in *Staphylococcus aureus*. *International Journal of Antimicrobial Agents*, *22*(3), 254–261. doi:10.1016/S0924-8579(03)00220-6 PMID:13678830

Kaatz, G. W., Moudgal, V. V., Seo, S. M., & Kristiansen, J. E. (2003). Phenothiazines and thioxanthenes inhibit multidrug efflux pump activity in *Staphylococcus aureus*. *Antimicrobial Agents and Chemotherapy*, *47*(2), 719–726. doi:10.1128/AAC.47.2.719-726.2003 PMID:12543683

Kalia, N. P., Mahajan, P., Mehra, R., Nargotra, A., Sharma, J. P., Koul, S., & Khan, I. A. (2012). Capsaicin, a novel inhibitor of the NorA efflux pump, reduces the intracellular invasion of Staphylococcus aureus. *The Journal of Antimicrobial Chemotherapy*, *67*(10), 2401–2408. doi:10.1093/jac/dks232 PMID:22807321

Kern, W. V., Steinke, P., Schumacher, A., Schuster, S., von Baum, H., & Bohnert, J. A. (2006). Effect of 1-(1-naphthylmethyl)-piperazine, a novel putative efflux pump inhibitor, on antimicrobial drug susceptibility in clinical isolates of Escherichia coli. *The Journal of Antimicrobial Chemotherapy*, *57*(2), 339–343. doi:10.1093/jac/dki445 PMID:16354747

Khan, I. A., Mirza, Z. M., Kumar, A., Verma, V., & Qazi, G. N. (2006). Piperine, a phytochemical potentiator of ciprofloxacin against Staphylococcus aureus. *Antimicrobial Agents and Chemotherapy*, *50*(2), 810–812. doi:10.1128/AAC.50.2.810-812.2006 PMID:16436753

Kikukawa, T., Nara, T., Araiso, T., Miyauchi, S., & Kamo, N. (2006). Two-component bacterial multidrug transporter, EbrAB: Mutations making each component solely functional. *Biomembranes*, *1758*(5), 673–679. doi:10.1016/j.bbamem.2006.04.004

Kiralj, R., & Ferreira, M. (2008). Comparative chemometric and QSAR/SAR study of structurally unrelated substrates of a MATE efflux pump VmrA from V. parahaemolyticus: Prediction of multidrug resistance. *QSAR & Combinatorial Science*, *27*(3), 314–329. doi:10.1002/qsar.200630164

Kolaczkowski, M., Michalak, K., & Motohashi, N. (2003). Phenothiazines as potent modulators of yeast multidrug resistance. *International Journal of Antimicrobial Agents*, *22*(3), 279–283. doi:10.1016/S0924-8579(03)00214-0 PMID:13678835

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Koronakis, V., Sharff, A., Koronakis, E., Luisi, B., & Hughes, C. (2000). Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. *Nature*, *405*(6789), 914–919. doi:10.1038/35016007 PMID:10879525

Koul, S., Koul, J. L., Taneja, S. C., Dhar, K. L., Jamwal, D. S., & Singh, K. et al. (2000). Structure–activity relationship of piperine and its synthetic analogues for their inhibitory potentials of rat hepatic microsomal constitutive and inducible cytochrome P450 activities. *Bioorganic & Medicinal Chemistry*, *8*(1), 251–268. doi:10.1016/S0968-0896(99)00273-4 PMID:10968285

Kumar, A., Khan, I. A., Koul, S., Koul, J. L., Taneja, S. C., & Ali, I. et al. (2008). Novel structural analogues of piperine as inhibitors of the NorA efflux pump of *Staphylococcus aureus. The Journal of Antimicrobial Chemotherapy*, *61*(6), 1270–1276. doi:10.1093/jac/dkn088 PMID:18334493

Kumar, S., & Varela, M. F. (2012). Biochemistry of bacterial multidrug efflux pumps. *International Journal of Molecular Sciences*, *13*(4), 4484–4495. doi:10.3390/ijms13044484 PMID:22605991

Lee, A., Mao, W., Warren, M. S., Mistry, A., Hoshino, K., & Okumura, R. et al. (2000). Interplay between efflux pumps may provide either additive or multiplicative effects on drug resistance. *Journal of Bacteriology*, *182*(11), 3142–3150. doi:10.1128/JB.182.11.3142-3150.2000 PMID:10809693

Lehtinen, J., & Lilius, E.-M. (2007). Promethazine renders *Escherichia coli* susceptible to penicillin G: Real-time measurement of bacterial susceptibility by fluoro-luminometry. *International Journal of Antimicrobial Agents*, *30*(1), 44–51. doi:10.1016/j.ijantimicag.2007.02.019 PMID:17475447

Lei, H.-T., Chou, T.-H., Su, C.-C., Bolla, J. R., Kumar, N., & Radhakrishnan, A. et al. (2014). Crystal Structure of the Open State of the *Neisseria gonorrhoeae* MtrE Outer Membrane Channel. *PLoS ONE*, *9*(6), e97475. doi:10.1371/journal.pone.0097475 PMID:24901251

Lei, H. T., Shen, Z., Surana, P., Routh, M. D., Su, C. C., Zhang, Q., & Yu, E. W. (2011). Crystal structures of CmeR-bile acid complexes from Campylobacter jejuni. *Protein Science*, *20*(4), 712–723. doi:10.1002/pro.602 PMID:21328631

Li, X.-Z., & Nikaido, H. (2009). Efflux-mediated drug resistance in bacteria. *Drugs*, *69*(12), 1555–1623. doi:10.2165/11317030-000000000-00000 PMID:19678712

Lin, J., Michel, L. O., & Zhang, Q. (2002). CmeABC functions as a multidrug efflux system in Campylobacter jejuni. *Antimicrobial Agents and Chemotherapy*, *46*(7), 2124–2131. doi:10.1128/AAC.46.7.2124-2131.2002 PMID:12069964

Lomovskaya, O., Warren, M. S., Lee, A., Galazzo, J., Fronko, R., & Lee, M. et al. (2001). Identification and characterization of inhibitors of multidrug resistance efflux pumps in *Pseudomonas aeruginosa*: Novel agents for combination therapy. *Antimicrobial Agents and Chemotherapy*, *45*(1), 105–116. doi:10.1128/AAC.45.1.105-116.2001 PMID:11120952

Lomovskaya, O., & Watkins, W. (2001). Inhibition of efflux pumps as a novel approach to combat drug resistance in bacteria. *Journal of Molecular Microbiology and Biotechnology*, *3*(2), 225–236. PMID:11321578

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Lu, M., Symersky, J., Radchenko, M., Koide, A., Guo, Y., Nie, R., & Koide, S. (2013). Structures of a Na+-coupled, substrate-bound MATE multidrug transporter. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(6), 2099–2104. doi:10.1073/pnas.1219901110 PMID:23341609

Lubelski, J., Konings, W. N., & Driessen, A. J. (2007). Distribution and physiology of ABC-type transporters contributing to multidrug resistance in bacteria. *Microbiology and Molecular Biology Reviews*, *71*(3), 463–476. doi:10.1128/MMBR.00001-07 PMID:17804667

Mahamoud, A., Chevalier, J., Alibert-Franco, S., Kern, W. V., & Pagès, J.-M. (2007). Antibiotic efflux pumps in Gram-negative bacteria: The inhibitor response strategy. *The Journal of Antimicrobial Chemotherapy*, *59*(6), 1223–1229. doi:10.1093/jac/dkl493 PMID:17229832

Malléa, M., Chevalier, J., Eyraud, A., & Pagès, J.-M. (2002). Inhibitors of antibiotic efflux pump in resistant *Enterobacter aerogenes* strains. *Biochemical and Biophysical Research Communications*, *293*(5), 1370–1373. doi:10.1016/S0006-291X(02)00404-7 PMID:12054665

Mannhold, R., Kubinyi, H., Folkers, G., Langer, T., & Hoffmann, R. D. (2006). *Pharmacophores and pharmacophore searches*. John Wiley & Sons.

Markham, P. N., & Neyfakh, A. A. (1996). Inhibition of the multidrug transporter NorA prevents emergence of norfloxacin resistance in Staphylococcus aureus. *Antimicrobial Agents and Chemotherapy*, *40*(11), 2673. PMID:8913490

Markham, P. N., & Neyfakh, A. A. (2001). Efflux-mediated drug resistance in Gram-positive bacteria. *Current Opinion in Microbiology*, *4*(5), 509–514. doi:10.1016/S1369-5274(00)00243-5 PMID:11587925

Mazzariol, A., Tokue, Y., Kanegawa, T. M., Cornaglia, G., & Nikaido, H. (2000). High-level fluoroquinolone-resistant clinical isolates of Escherichia coli overproduce multidrug efflux protein AcrA. *Antimicrobial Agents and Chemotherapy*, *44*(12), 3441–3443. doi:10.1128/AAC.44.12.3441-3443.2000 PMID:11083655

McMurry, L., Petrucci, R. E., & Levy, S. B. (1980). Active efflux of tetracycline encoded by four genetically different tetracycline resistance determinants in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, *77*(7), 3974–3977. doi:10.1073/pnas.77.7.3974 PMID:7001450

Meng, J., Bai, H., Jia, M., Ma, X., Hou, Z., & Xue, X. et al. (2011). Restoration of antibiotic susceptibility in fluoroquinolone-resistant Escherichia coli by targeting acrB with antisense phosphorothioate oligonucleotide encapsulated in novel anion liposome. *The Journal of Antibiotics*, *65*(3), 129–134. doi:10.1038/ja.2011.125 PMID:22186597

Michalak, K., Wesolowska, O., Motohashi, N., Molnar, J., & Hendrich, A. (2006). Interactions of phenothiazines with lipid bilayer and their role in multidrug resistance reversal. *Current Drug Targets*, *7*(9), 1095–1105. doi:10.2174/138945006778226570 PMID:17017888

Michalet, S., Cartier, G., David, B., Mariotte, A.-M., Dijoux-franca, M.-G., & Kaatz, G. W. et al. (2007). *N*-Caffeoylphenalkylamide derivatives as bacterial efflux pump inhibitors. *Bioorganic & Medicinal Chemistry Letters*, *17*(6), 1755–1758. doi:10.1016/j.bmcl.2006.12.059 PMID:17275293

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Molnar, J., Hever, A., Fakla, I., Fischer, J., Ocsovski, I., & Aszalos, A. (1996). Inhibition of the transport function of membrane proteins by some substituted phenothiazines in E. coli and multidrug resistant tumor cells. *Anticancer Research*, *17*(1A), 481–486. PMID:9066699

Munoz-Bellido, J., Munoz-Criado, S., & Garcìa-Rodrìguez, J. (2000). Antimicrobial activity of psychotropic drugs: Selective serotonin reuptake inhibitors. *International Journal of Antimicrobial Agents*, *14*(3), 177–180. doi:10.1016/S0924-8579(99)00154-5 PMID:10773485

Murakami, S., Nakashima, R., Yamashita, E., & Yamaguchi, A. (2002). Crystal structure of bacterial multidrug efflux transporter AcrB. *Nature*, *419*(6907), 587–593. doi:10.1038/nature01050 PMID:12374972

Nakashima, R., Sakurai, K., Yamasaki, S., Hayashi, K., Nagata, C., & Hoshino, K. et al. (2013). Structural basis for the inhibition of bacterial multidrug exporters. *Nature*, *500*(7460), 102–106. doi:10.1038/nature12300 PMID:23812586

Nakayama, K., Kawato, H., Watanabe, J., Ohtsuka, M., Yoshida, K.-i., & Yokomizo, Y. et al. (2004). MexAB-OprM specific efflux pump inhibitors in *Pseudomonas aeruginosa*. Part 3: Optimization of potency in the pyridopyrimidine series through the application of a pharmacophore model. *Bioorganic & Medicinal Chemistry Letters*, *14*(2), 475–479. doi:10.1016/j.bmcl.2003.10.060 PMID:14698185

Nargotra, A., Koul, S., Sharma, S., Khan, I., Kumar, A., & Thota, N. et al. (2009). Quantitative structure–activity relationship (QSAR) of aryl alkenyl amides/imines for bacterial efflux pump inhibitors. *European Journal of Medicinal Chemistry*, *44*(1), 229–238. doi:10.1016/j.ejmech.2008.02.015 PMID:18395941

Nargotra, A., Sharma, S., Koul, J. L., Sangwan, P. L., Khan, I. A., & Kumar, A. et al. (2009). Quantitative structure activity relationship (QSAR) of piperine analogs for bacterial NorA efflux pump inhibitors. *European Journal of Medicinal Chemistry*, *44*(10), 4128–4135. doi:10.1016/j.ejmech.2009.05.004 PMID:19523722

Nargotra, A., Sharma, S., Koul, J. L., Sangwan, P. L., Khan, I. A., & Kumar, A. et al. (2009). Quantitative structure activity relationship (QSAR) of piperine analogsfor bacterial NorA efflux pump inhibitors. *European Journal of Medicinal Chemistry*, *44*(10), 4128–4135. doi:10.1016/j.ejmech.2009.05.004 PMID:19523722

Neu, H. C. (1992). The crisis in antibiotic resistance. *Science*, *257*(5073), 1064–1073. doi:10.1126/science.257.5073.1064 PMID:1509257

Nikaido, H. (1996). Multidrug efflux pumps of gram-negative bacteria. *Journal of Bacteriology*, *178*(20), 5853. PMID:8830678

Nikaido, H. (2001). *Preventing drug access to targets: cell surface permeability barriers and active efflux in bacteria.* Paper presented at the Seminars in Cell & Developmental Biology. doi:10.1006/scdb.2000.0247

Nikaido, H. (2003). Molecular basis of bacterial outer membrane permeability revisited. *Microbiology and Molecular Biology Reviews*, *67*(4), 593–656. doi:10.1128/MMBR.67.4.593-656.2003 PMID:14665678

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Nikaido, H., Basina, M., Nguyen, V., & Rosenberg, E. Y. (1998). Multidrug efflux pump acrab of *salmonella typhimurium* excretes only those β-lactam antibiotics containing lipophilic side chains. *Journal of Bacteriology*, *180*(17), 4686–4692. PMID:9721312

Oethinger, M., & Levy, S. B. (2002). *CA patent no. 2378883*. C. I. P. Office.

Ojha, P. K., Mitra, I., Das, R. N., & Roy, K. (2011). Further exploring rm^2 metrics for validation of QSPR models. *Chemometrics and Intelligent Laboratory Systems*, *107*(1), 194–205. doi:10.1016/j.chemolab.2011.03.011

Omote, H., Hiasa, M., Matsumoto, T., Otsuka, M., & Moriyama, Y. (2006). The MATE proteins as fundamental transporters of metabolic and xenobiotic organic cations. *Trends in Pharmacological Sciences*, *27*(11), 587–593. doi:10.1016/j.tips.2006.09.001 PMID:16996621

Pagès, J.-M., Masi, M., & Barbe, J. (2005). Inhibitors of efflux pumps in gram-negative bacteria. *Trends in Molecular Medicine*, *11*(8), 382–389. doi:10.1016/j.molmed.2005.06.006 PMID:15996519

Park, Y., Lee, Y., Kim, H., Lee, Y., Yoon, Y., & Moon, B. et al. (2006). NMR data of flavone derivatives and their anti-oxidative activities. *Bulletin of the Korean Chemical Society*, *27*(10), 1537–1541. doi:10.5012/bkcs.2006.27.10.1537

Paulsen, I. T., Brown, M. H., & Skurray, R. A. (1996). Proton-dependent multidrug efflux systems. *Microbiological Reviews*, *60*(4), 575–608. PMID:8987357

Phan, G., Benabdelhak, H., Lascombe, M.-B., Benas, P., Rety, S., & Picard, M. et al. (2010). Structural and dynamical insights into the opening mechanism of *p. aeruginosa* OprM channel. *Structure (London, England)*, *18*(4), 507–517. doi:10.1016/j.str.2010.01.018 PMID:20399187

Phosrithong, N., Samee, W., Nunthanavanit, P., & Ungwitayatorn, J. (2012). *In vitro* antioxidant activity study of novel chromone derivatives. *Chemical Biology & Drug Design*, *79*(6), 981–989. doi:10.1111/j.1747-0285.2012.01368.x PMID:22381130

Pieroni, M., Dimovska, M., Brincat, J. P., Sabatini, S., Carosati, E., & Massari, S. et al. (2010). From 6-aminoquinolone antibacterials to 6-amino-7-thiopyranopyridinylquinolone ethyl esters as inhibitors of *Staphylococcus aureus* multidrug efflux pumps. *Journal of Medicinal Chemistry*, *53*(11), 4466–4480. doi:10.1021/jm1003304 PMID:20446747

Poole, K. (2000). Efflux-mediated resistance to fluoroquinolones in gram-negative bacteria. *Antimicrobial Agents and Chemotherapy*, *44*(9), 2233–2241. doi:10.1128/AAC.44.9.2233-2241.2000 PMID:10952561

Poole, K. (2005). Efflux-mediated antimicrobial resistance. *The Journal of Antimicrobial Chemotherapy*, *56*(1), 20–51. doi:10.1093/jac/dki171 PMID:15914491

Pos, K. M., & Diederichs, K. (2002). Purification, crystallization and preliminary diffraction studies of AcrB, an inner-membrane multi-drug efflux protein. *Acta Crystallographica. Section D, Biological Crystallography*, *58*(10), 1865–1867. doi:10.1107/S0907444902013963 PMID:12351840

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Pratim Roy, P., Paul, S., Mitra, I., & Roy, K. (2009). On two novel parameters for validation of predictive QSAR models. *Molecules (Basel, Switzerland)*, *14*(5), 1660–1701. doi:10.3390/molecules14051660 PMID:19471190

Pumbwe, L., & Piddock, L. J. (2002). Identification and molecular characterisation of CmeB, a *Campylobacter jejuni* multidrug efflux pump. *FEMS Microbiology Letters*, *206*(2), 185–189. doi:10.1111/j.1574-6968.2002.tb11007.x PMID:11814661

Putman, M., van Veen, H. W., & Konings, W. N. (2000). Molecular properties of bacterial multidrug transporters. *Microbiology and Molecular Biology Reviews*, *64*(4), 672–693. doi:10.1128/MMBR.64.4.672-693.2000 PMID:11104814

Renau, T. E., Léger, R., Flamme, E. M., Sangalang, J., She, M. W., & Yen, R. et al. (1999). Inhibitors of efflux pumps in *Pseudomonas aeruginosa* potentiate the activity of the fluoroquinolone antibacterial levofloxacin. *Journal of Medicinal Chemistry*, *42*(24), 4928–4931. doi:10.1021/jm9904598 PMID:10585202

Rosenbusch, J. P. (2001). Stability of membrane proteins: Relevance for the selection of appropriate methods for high-resolution structure determinations. *Journal of Structural Biology*, *136*(2), 144–157. doi:10.1006/jsbi.2001.4431 PMID:11886216

Roy, K., Mitra, I., Kar, S., Ojha, P. K., Das, R. N., & Kabir, H. (2011). Comparative studies on some metrics for external validation of QSPR models. *Journal of Chemical Information and Modeling*, *52*(2), 396–408. doi:10.1021/ci200520g PMID:22201416

Ruiz, J. (2003). Mechanisms of resistance to quinolones: Target alterations, decreased accumulation and DNA gyrase protection. *The Journal of Antimicrobial Chemotherapy*, *51*(5), 1109–1117. doi:10.1093/jac/dkg222 PMID:12697644

Sabatini, S., Gosetto, F., Serritella, S., Manfroni, G., Tabarrini, O., & Iraci, N. et al. (2012). Pyrazolo [4, 3-c][1, 2] benzothiazines 5, 5-dioxide: A promising new class of *Staphylococcus aureus* NorA efflux pump inhibitors. *Journal of Medicinal Chemistry*, *55*(7), 3568–3572. doi:10.1021/jm201446h PMID:22432682

Sabatini, S., Kaatz, G. W., Rossolini, G. M., Brandini, D., & Fravolini, A. (2008). From phenothiazine to 3-phenyl-1, 4-benzothiazine derivatives as inhibitors of the *Staphylococcus aureus* NorA multidrug efflux pump. *Journal of Medicinal Chemistry*, *51*(14), 4321–4330. doi:10.1021/jm701623q PMID:18578473

Saier, M. H. Jr. (1998). Molecular phylogeny as a basis for the classification of transport proteins from bacteria, archaea and eukarya. *Advances in Microbial Physiology*, *40*, 81–136. doi:10.1016/S0065-2911(08)60130-7 PMID:9889977

Saier, M. H. Jr, Beatty, J. T., Goffeau, A., Harley, K. T., Heijne, W., & Huang, S.-C. et al. (1999). The major facilitator superfamily. *Journal of Molecular Microbiology and Biotechnology*, *1*(2), 257–279. PMID:10943556

Schindler, B. D., Jacinto, P., & Kaatz, G. W. (2013). Inhibition of drug efflux pumps in *Staphylococcus aureus*: Current status of potentiating existing antibiotics. *Future Microbiology*, *8*(4), 491–507. doi:10.2217/fmb.13.16 PMID:23534361

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Seeger, M. A., Schiefner, A., Eicher, T., Verrey, F., Diederichs, K., & Pos, K. M. (2006). Structural asymmetry of AcrB trimer suggests a peristaltic pump mechanism. *Science*, *313*(5791), 1295–1298. doi:10.1126/science.1131542 PMID:16946072

Sennhauser, G., Bukowska, M. A., Briand, C., & Grütter, M. G. (2009). Crystal structure of the multidrug exporter MexB from *Pseudomonas aeruginosa. Journal of Molecular Biology*, *389*(1), 134–145. doi:10.1016/j.jmb.2009.04.001 PMID:19361527

Seyoum, A., Asres, K., & El-Fiky, F. K. (2006). Structure–radical scavenging activity relationships of flavonoids. *Phytochemistry*, *67*(18), 2058–2070. doi:10.1016/j.phytochem.2006.07.002 PMID:16919302

Singh, J., Dubey, R., & Atal, C. (1986). Piperine-mediated inhibition of glucuronidation activity in isolated epithelial cells of the guinea-pig small intestine: Evidence that piperine lowers the endogeneous UDP-glucuronic acid content. *The Journal of Pharmacology and Experimental Therapeutics*, *236*(2), 488–493. PMID:3080587

Smith, E. C., Williamson, E. M., Wareham, N., Kaatz, G. W., & Gibbons, S. (2007). Antibacterials and modulators of bacterial resistance from the immature cones of *Chamaecyparis lawsoniana. Phytochemistry*, *68*(2), 210–217. doi:10.1016/j.phytochem.2006.10.001 PMID:17109904

Stavri, M., Piddock, L. J., & Gibbons, S. (2007). Bacterial efflux pump inhibitors from natural sources. *The Journal of Antimicrobial Chemotherapy*, *59*(6), 1247–1260. doi:10.1093/jac/dkl460 PMID:17145734

Tegos, G. P., Haynes, M., Strouse, J. J., Khan, M. M. T., Bologa, C. G., Oprea, T. I., & Sklar, L. A. (2011). Microbial efflux pump inhibition: Tactics and strategies. *Current Pharmaceutical Design*, *17*(13), 1291. doi:10.2174/138161211795703726 PMID:21470111

Thai, K.-M., Ngo, T.-D., Phan, T.-V., Tran, T.-D., Nguyen, N.-V., Nguyen, T.-H., & Le, M.-T. (2015). Virtual screening for novel *Staphylococcus aureus* NorA efflux pump inhibitors from natural products. *Medicinal Chemistry* 11(2), 135-155. doi: 10.2174/1573406410666140902110903.

Thanabalu, T., Koronakis, E., Hughes, C., & Koronakis, V. (1998). Substrate-induced assembly of a contiguous channel for protein export from *E. coli*: Reversible bridging of an inner-membrane translocase to an outer membrane exit pore. *The EMBO Journal*, *17*(22), 6487–6496. doi:10.1093/emboj/17.22.6487 PMID:9822594

Thorarensen, A., Presley-Bodnar, A. L., Marotti, K. R., Boyle, T. P., Heckaman, C. L., & Bohanon, M. J. et al. (2001). 3-Arylpiperidines as potentiators of existing antibacterial agents. *Bioorganic & Medicinal Chemistry Letters*, *11*(14), 1903–1906. doi:10.1016/S0960-894X(01)00330-4 PMID:11459657

Tran, T.-D., Park, H., Kim, H. P., Ecker, G. F., & Thai, K.-M. (2009). Inhibitory activity of prostaglandin E2 production by the synthetic 2′-hydroxychalcone analogues: Synthesis and SAR study. *Bioorganic & Medicinal Chemistry Letters*, *19*(6), 1650–1653. doi:10.1016/j.bmcl.2009.02.001 PMID:19233646

Tseng, T.-T., Gratwick, K. S., Kollman, J., Park, D., Nies, D. H., Goffeau, A., & Saier, M. H. Jr. (1999). The RND permease superfamily: An ancient, ubiquitous and diverse family that includes human disease and development proteins. *Journal of Molecular Microbiology and Biotechnology*, *1*(1), 107–125. PMID:10941792

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Van Bambeke, F., Balzi, E., & Tulkens, P. M. (2000). Antibiotic efflux pumps. *Biochemical Pharmacology*, *60*(4), 457–470. doi:10.1016/S0006-2952(00)00291-4 PMID:10874120

Van Bambeke, F., & Lee, V. J. (2006). Inhibitors of bacterial efflux pumps as adjuvants in antibiotic treatments and diagnostic tools for detection of resistance by efflux. *Recent Patents on Anti-infective Drug Discovery*, *1*(2), 157–175. doi:10.2174/157489106777452692 PMID:18221142

Vargiu, A. V., & Nikaido, H. (2012). Multidrug binding properties of the AcrB efflux pump characterized by molecular dynamics simulations. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(50), 20637–20642. doi:10.1073/pnas.1218348109 PMID:23175790

Wang, K., Pei, H., Huang, B., Zhu, X., Zhang, J., & Zhou, B. et al. (2013). The expression of ABC efflux pump, Rv1217c–Rv1218c, and its association with multidrug resistance of *Mycobacterium tuberculosis* in China. *Current Microbiology*, *66*(3), 222–226. doi:10.1007/s00284-012-0215-3 PMID:23143285

Wasaznik, A., Grinholc, M., & Bielawski, K. P. (2008). Active efflux as the multidrug resistance mechanism. *Postepy higieny i medycyny doswiadczalnej (Online)*, *63*, 123–133. PMID:19373193

Wood, A. J., Gold, H. S., & Moellering, R. C. Jr. (1996). Antimicrobial-drug resistance. *The New England Journal of Medicine*, *335*(19), 1445–1453. doi:10.1056/NEJM199611073351907 PMID:8875923

Yilmaz, S., Altinkanat-Gelmez, G., Bolelli, K., Guneser-Merdan, D., Over-Hasdemir, M. U., & Yildiz, I. et al. (2014). Pharmacophore generation of 2-substituted benzothiazoles as AdeABC efflux pump inhibitors in *A. baumannii. SAR and QSAR in Environmental Research*, *25*(7), 551–563. doi:10.1080/1062936X.2014.919357 PMID:24905472

Yokozawa, T., Chen, C. P., Dong, E., Tanaka, T., Nonaka, G.-I., & Nishioka, I. (1998). Study on the inhibitory effect of tannins and flavonoids against the 1,1-diphenyl-2-picrylhydrazyl radical. *Biochemical Pharmacology*, *56*(2), 213–222. doi:10.1016/S0006-2952(98)00128-2 PMID:9698075

Yum, S., Xu, Y., Piao, S., Sim, S.-H., Kim, H.-M., & Jo, W.-S. et al. (2009). Crystal structure of the periplasmic component of a tripartite macrolide-specific efflux pump. *Journal of Molecular Biology*, *387*(5), 1286–1297. doi:10.1016/j.jmb.2009.02.048 PMID:19254725

Zechini, B., & Versace, I. (2009). Inhibitors of multidrug resistant efflux systems in bacteria. *Recent Patents on Anti-infective Drug Discovery*, *4*(1), 37–50. doi:10.2174/157489109787236256 PMID:19149695

Zhang, L., & Ma, S. (2010). Efflux pump inhibitors: A strategy to combat p-glycoprotein and the NorA multidrug resistance pump. *ChemMedChem*, *5*(6), 811–822. doi:10.1002/cmdc.201000006 PMID:20373322

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

KEY TERMS AND DEFINITIONS

Bacterial Efflux Pumps: Proteinaceous transporters located in the cytoplasmic membrane of bacteria. These transporters are responsible for moving a wide range of compounds including antibiotics out of bacterial cells, which can contribute to bacterial antibiotic resistance.

Docking Study: The method used for the simulation of optimal binding orientations of one molecule to its target protein thereby predicting the affinity and biological activity of small organic molecules. Docking may be applied in hit identification and lead optimization in the field of drug design.

Efflux Pump Inhibitors (EPIs): Compounds that can inhibit the efflux-mediated extrusion of antibiotics. The co-administrations of these compounds and antibiotics to restore antibiotic activity have been investigated, but none of the efflux inhibitors tested has been in clinical use until now.

Homology Modeling (Also Known as Comparative Modeling): The method used for construction of an atomic-resolution model of the target protein based on its amino acid sequence and an empirically determined 3D-structure of a template – the protein having the highest sequence identity with the target sequence. A homology model can be used for various purposes including molecular docking, protein-protein interaction prediction, homology model based ligand design, etc.

MultiDrug Resistance (MDR): A condition enabling pathogenic micro-organisms (including bacteria) to oppose the action of certain antibiotics, rendering the antibiotics ineffective.

Pharmacophore Hypothesis: A geometrical description of molecular features represents a binding mode of a ligand to a receptor site. Typical pharmacophore features include hydrophobic centers, aromatic rings, ionizable points, hydrogen bond donors and acceptors.

Protein Data Bank: A freely accessible source for 3D-structures of biological macromolecules such as proteins and nucleic acids. Most of these structures are determined by NMR spectroscopy and X-ray diffraction.

Protein X-Ray Crystallography: The method of characterizing the molecular structure and function of various biological molecules including proteins. In principle, when X-rays strikes a protein crystal, they are scattered into many specific directions. The patterns and relative intensities of these diffracted X-ray beams can be processed to provide information about protein structure.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 8
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: 10.4018/978-1-4666-8136-1.ch008

Copyright © 2015, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/14/2023 7:16 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.

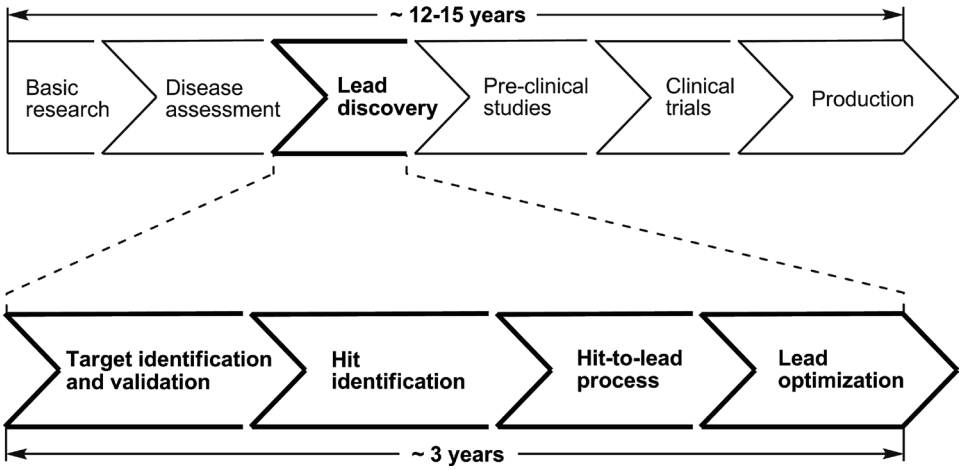
As depicted in Figure 1, the lead discovery part (which is constructed of several interconnected sub-phases) of the drug discovery pipeline can be considered as the essential one (Kenny et al., 1998; Langer & Hoffmann, 2001). Owing to the rapid development of various computational methods that today 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 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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 1. Schematic representation of a classical industrial drug discovery pipeline



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 a major challenge of many computational chemists (Salemme et al., 1997; Lewis, 2005; Nevin et al., 2012; Tian et al., 2013; Chen, 2013).

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

EBSCOhost - printed on 2/14/2023 7:16 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

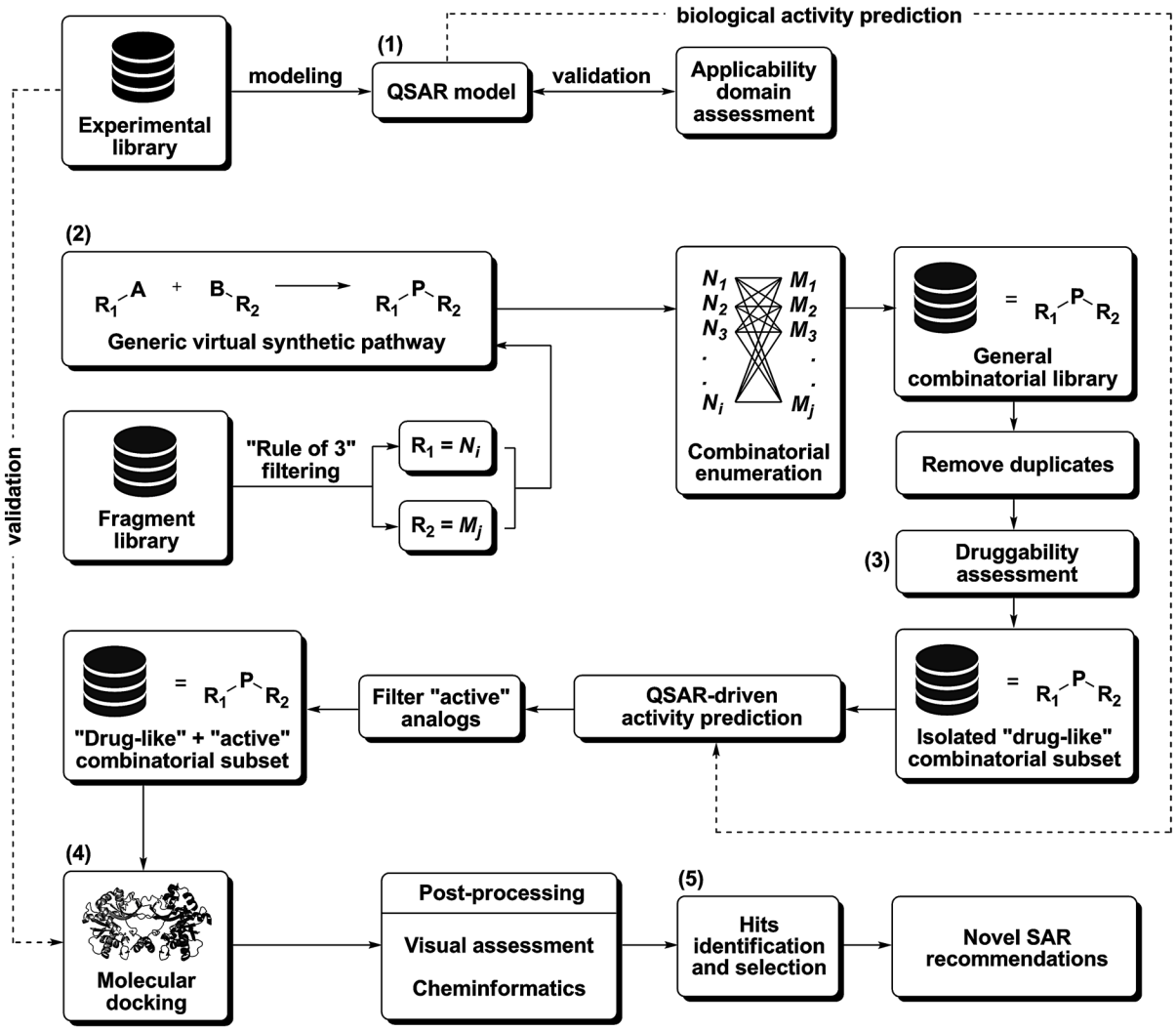
pre-processing strategies (e.g., QSAR-based activity predictions and screening, "drug-likeness" filtering, etc.). Numerous examples demonstrating a high successful rate when using such combined approaches are available elsewhere in the literature and some of the recent are summarized in Table 1.

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 2/14/2023 7:16 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 2/14/2023 7:16 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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

al., 2007). Regardless of its wide application for plain numerical predictions of biological activities or 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).

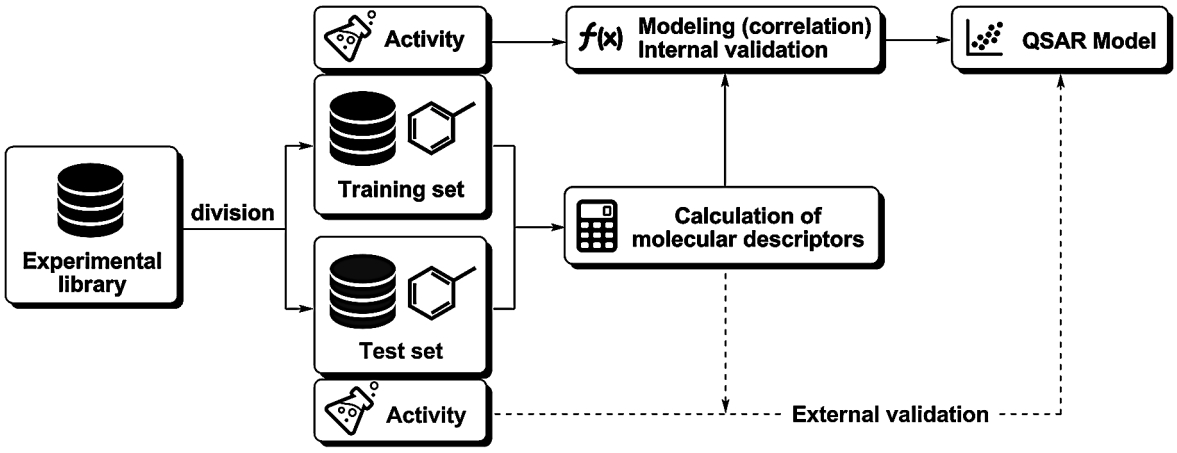
As shown in Figure 3, only training set compounds are considered in the construction of the QSAR model. In a mathematical sense, the aim is to establish a correlation between the chemical structures of the investigated compounds (training set objects) expressed in a form of calculated molecular descriptors and their corresponding biological activity values (Equation 1).

$$pA = c_0 + c_1 d_1 + c_2 d_2 + c_3 d_3 + ... + c_i d_i \quad (1)$$

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 3. A general QSAR methodology flowchart



(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).

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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.

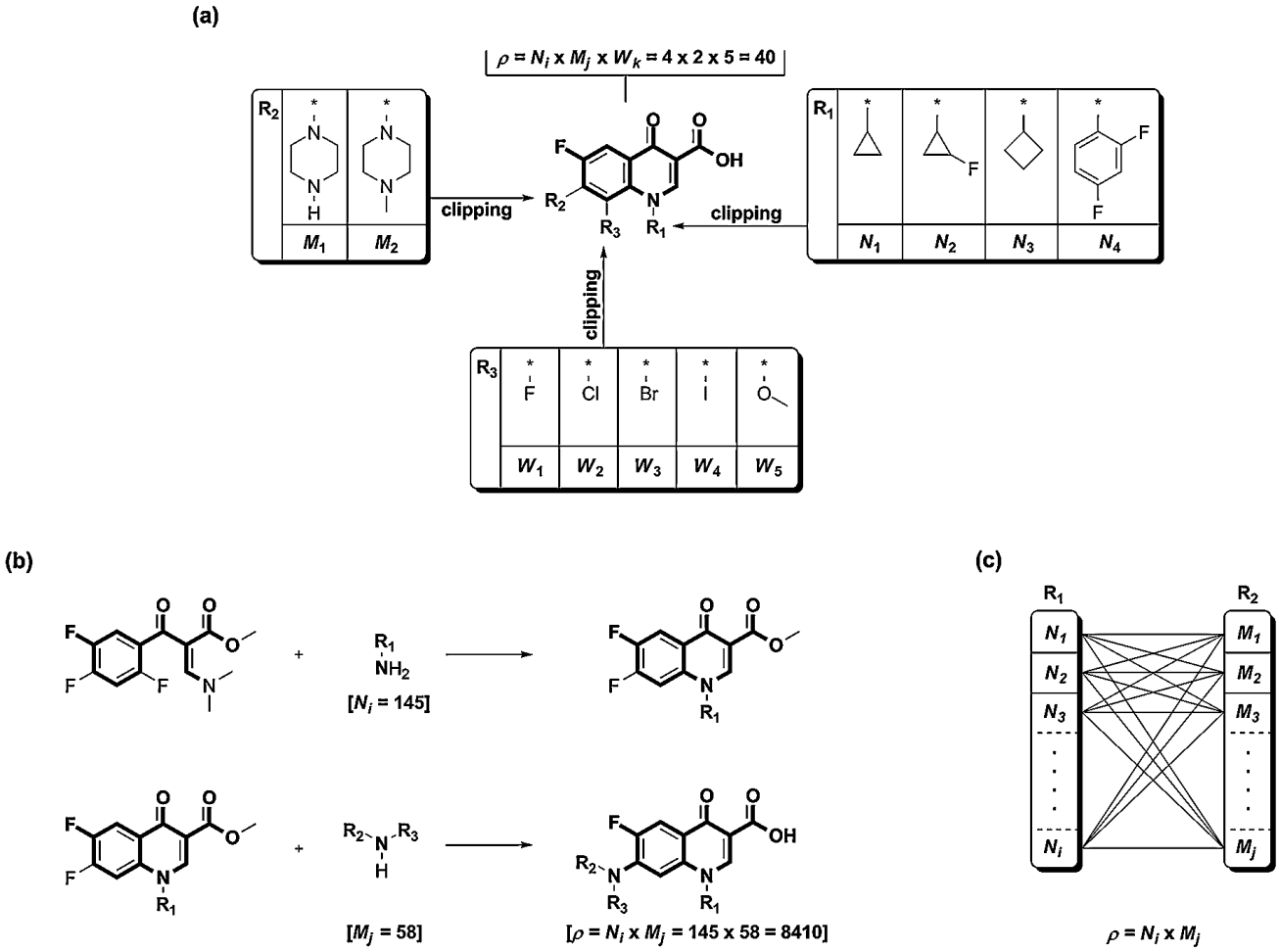
Two generally accepted strategies are commonly used for virtual enumeration of combinatorial libraries such as the Markush enumeration approach (Leland et al., 1997) as well as the reaction-based enumeration approach (Leach et al., 1999; Lobanov & Agrafiotis, 2002). In the Markush enumeration, the combinatorial 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

EBSCOhost - printed on 2/14/2023 7:16 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)



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., 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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 ($N_1 \ldots N_i$) and R_2 ($M_1 \ldots M_j$). 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 ($N_1 \ldots N_i$ and $M_1 \ldots M_j$). Mathematically, this process can be expressed by using the following equation (Equation 4):

$$\rho = N_i \times M_j \tag{4}$$

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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).

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 $\leq$ 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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 $0.0 \leq \text{MIC}\ [\mu g/mL] \leq 1.0\text{-}2.0$, 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).

Structure-Based Virtual Screening and Post-Processing of QSAR Prioritized "Drug-Like" Combinatorial Compounds

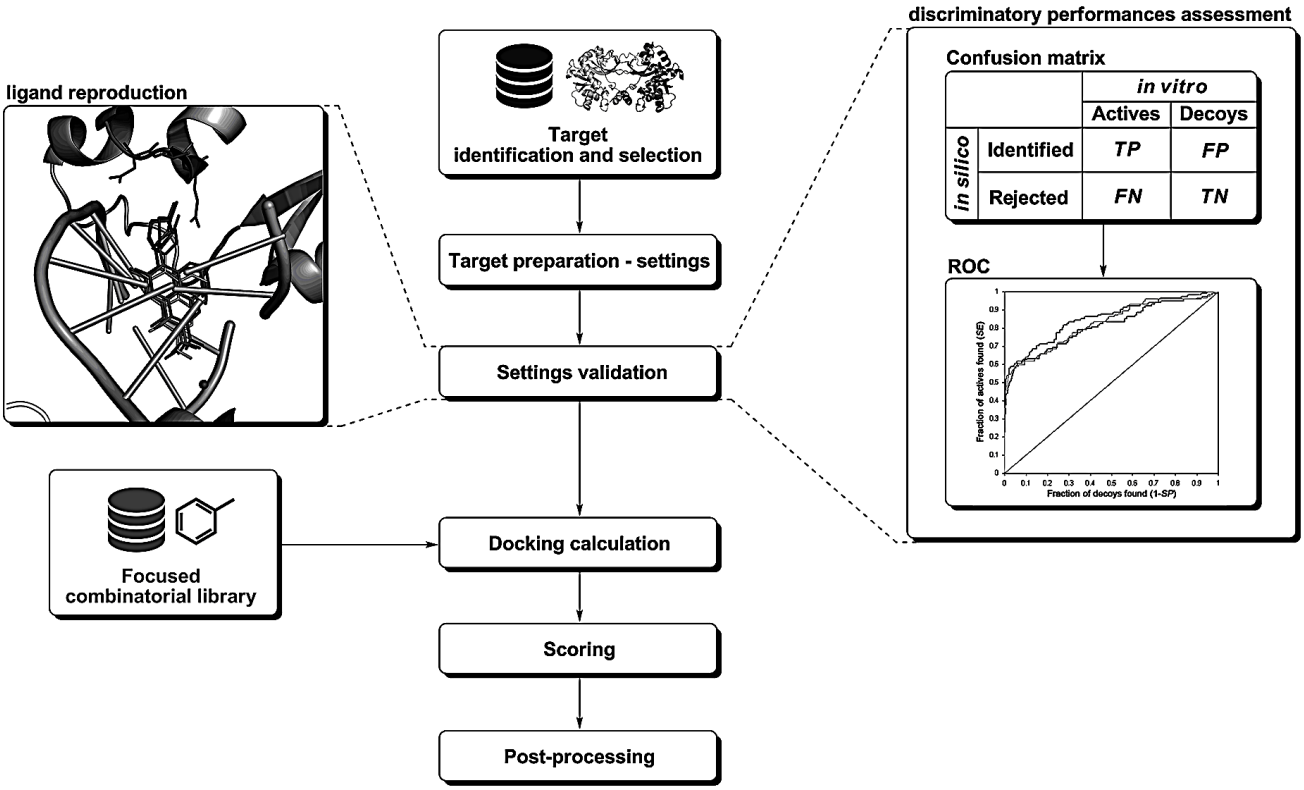
The *in silico* strategies elaborated so far generally belong to the class of ligand-based methodologies which usefulness and versatility in various drug discovery segments are broadly reviewed (Vidal et al., 2011). Notwithstanding their mounting success, a major drawback that follows all these approaches 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

EBSCOhost - printed on 2/14/2023 7:16 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



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 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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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):

$$SE = \frac{TP}{TP + FN} \tag{5}$$

$$SP = \frac{TN}{TN + FP} \tag{6}$$

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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, 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)

$$\text{EF}_{x\%} = \frac{\dfrac{N^*_A}{N_{x\%}}}{\dfrac{N_A}{N_A + N_D}} \tag{8}$$

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

(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 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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 quantitatively. 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 MIC ≤ 0.05 μg/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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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.

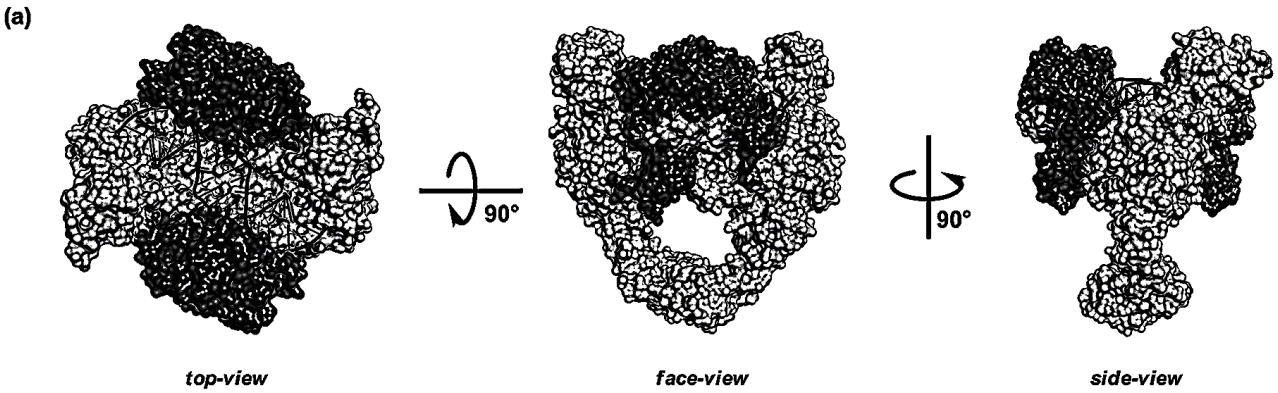
M. tuberculosis DNA Gyrase as a Therapeutic Target and 6-Fluoroquinolone Antibacterials

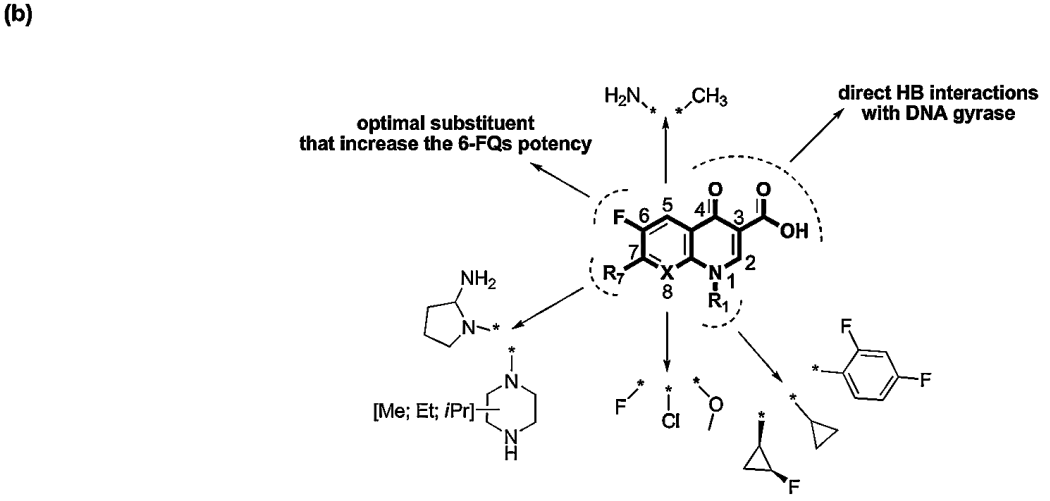
It is widely known that in all living organisms, the correct spatial DNA topology is of crucial importance for the proper regulation of the DNA processing including all the basic biological processes, control of the gene replication and transcription, as well as the DNA segregation (Wasserman & Cozzarelli 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).

EBSCOhost - printed on 2/14/2023 7:16 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





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).

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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.

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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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

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).

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

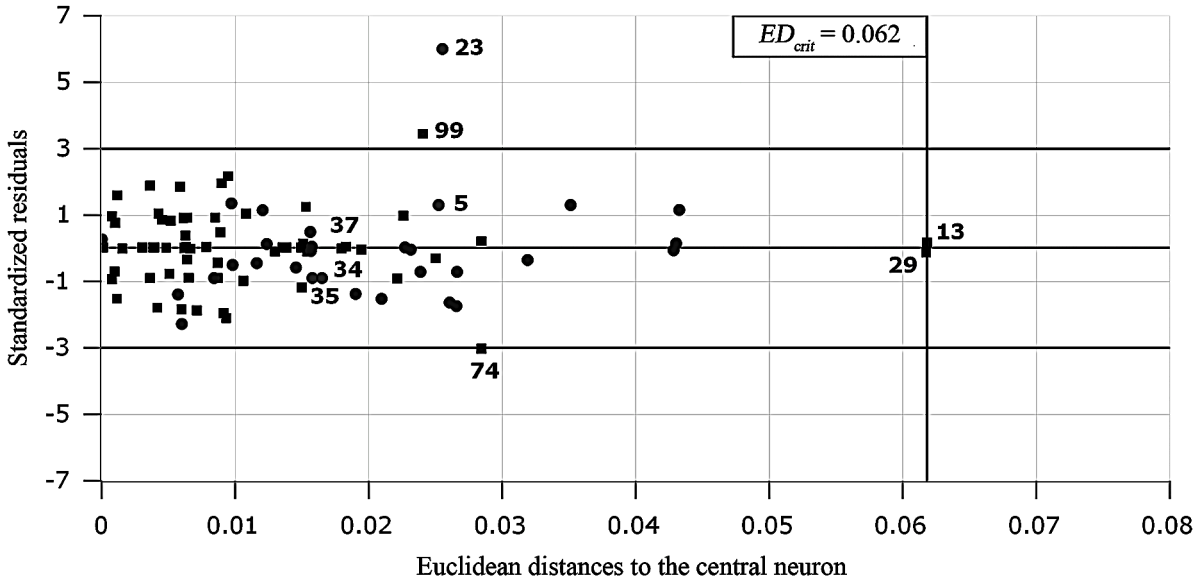
Following the integrated *in silico* screening protocol described previously (Figure 2), non-linear artificial neural networks (ANNs)-based predictive QSAR modeling was performed on the *ExpLib* library of 145 6-FQs using a comprehensive set of approximately 600 calculated 2D molecular descriptors (Minovski 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 procedure, 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 ($pMIC_{exp[ID=23]} = 0.7404$, $pMIC_{pred[ID=23]} = 0.2334$). The structural analysis

EBSCOhost - printed on 2/14/2023 7:16 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.



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 ±3σ 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 ($MIC_{pred\text{-}combi}$ [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 $0.0021 \leq MIC_{pred\text{-}combi}$ [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 $0.00 \leq MIC_{pred\text{-}combi}$ [μg/mL] ≤ 1.00, which served as

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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) 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 co-factors 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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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 ($ROC\text{-}AUC_{[3K9Fmod]} = 0.844$), $2XCT_{mod}$ has somehow moderate VS performances ($ROC\text{-}AUC_{[2XCTmod]} = 0.826$), while $2XKK_{mod}$ could be identified as a poorest one ($ROC\text{-}AUC_{[2XKKmod]} = 0.815$). 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 ($p\text{-value}_{[2XKKmod]} = 0.192$ and $p\text{-value}_{[2XCTmod]} = 0.299$, 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 ($p\text{-value}_{[2XCTmod]} > p\text{-value}_{[2XKKmod]}$), 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 (2XK-

EBSCOhost - printed on 2/14/2023 7:16 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).

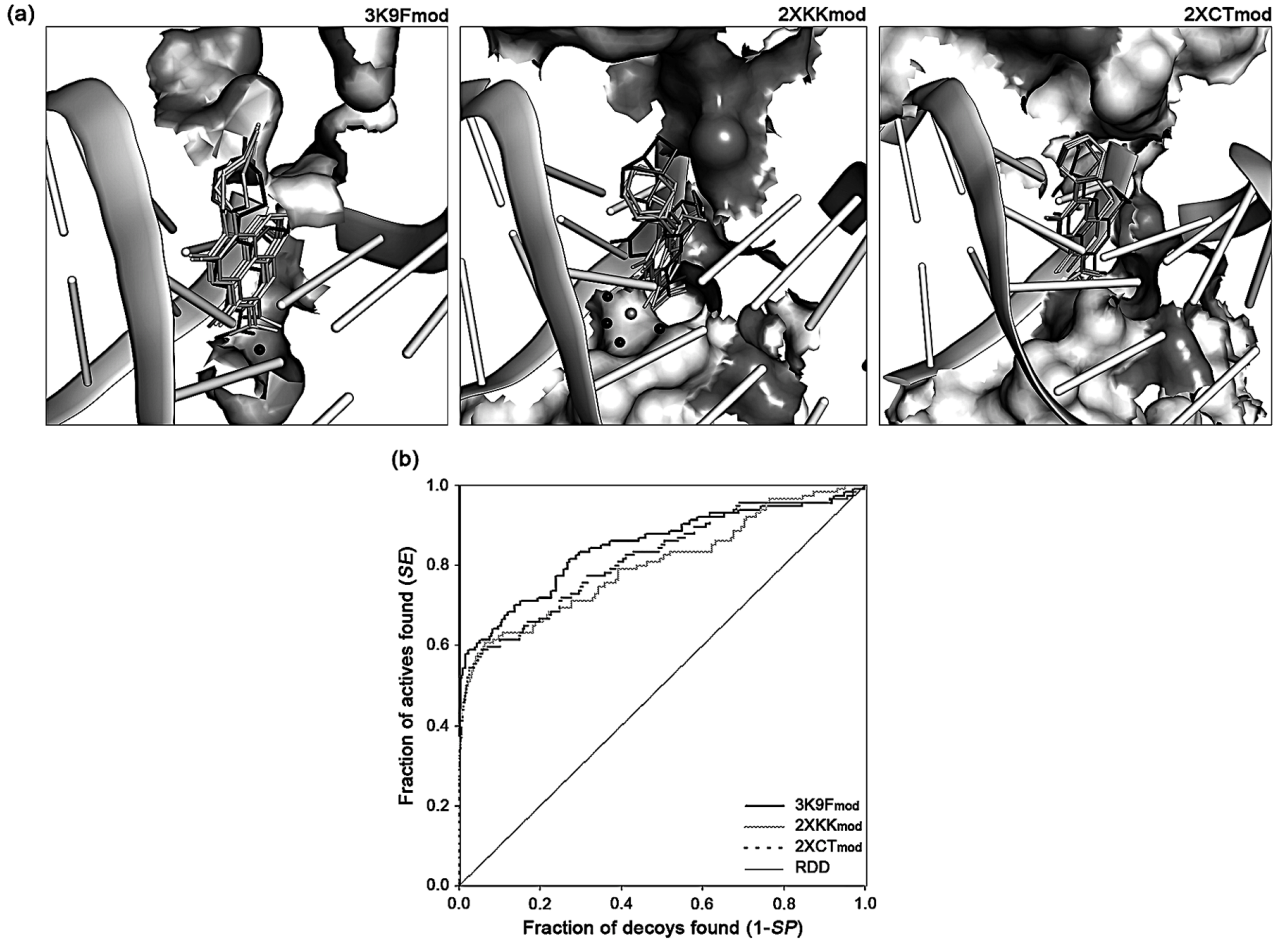


Table 3. Ligand reproduction assessment for initial validation of the quality of the assembled M. tuberculosis DNA gyrase protein homology models (3K9F$_{mod}$, 2XKK$_{mod}$, and 2XCT$_{mod}$, respectively) performed by heavy-atoms RMSD (Å) comparison between each experimental ligand conformation (LFX, 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 2/14/2023 7:16 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

$K_{mod} < 2XCT_{mod} < 3K9F_{mod}$). 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).

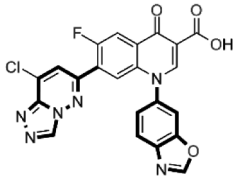
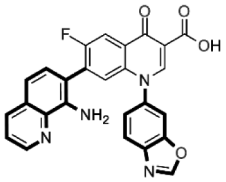
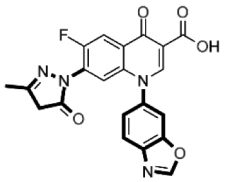
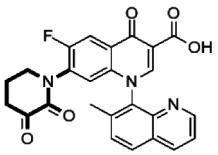
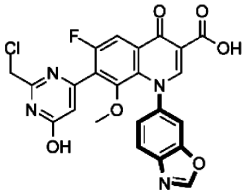
Boolean-Based Clustering for Identification and Selection of Novel 6-FQ Hits

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

EBSCOhost - printed on 2/14/2023 7:16 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

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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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.

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. 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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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.

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 2/14/2023 7:16 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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 9
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: 10.4018/978-1-4666-8136-1.ch009

Copyright © 2015, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/14/2023 7:16 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 2/14/2023 7:16 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:

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

1. Extraction of Molecular descriptors from structures,
2. Selection of relevant molecular descriptors, and
3. Generation of QSAR model. (Duke et al., 2006)

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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. 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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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.

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 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

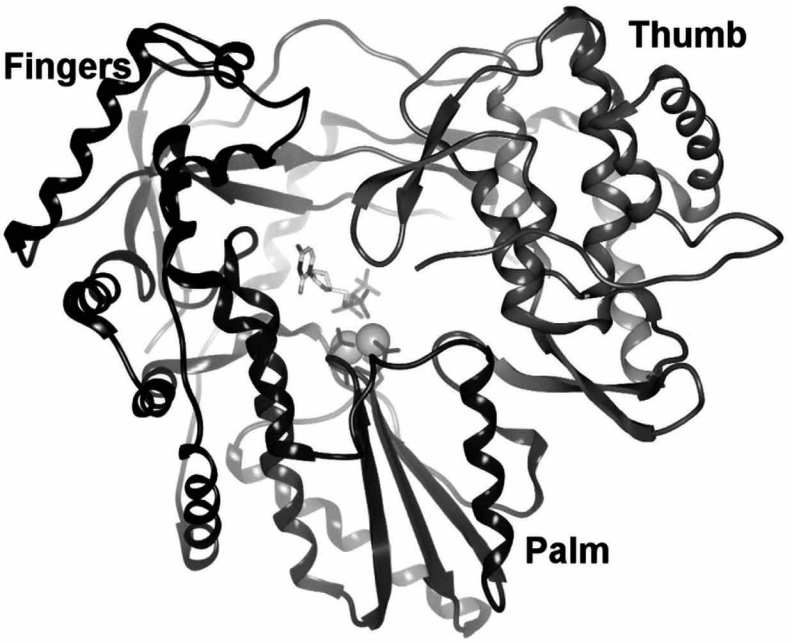
NONSTRUCTURAL PROTEIN 5B (NS5B)

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).

Figure 1. Structure of NS5B (PDB ID: 2DXS)



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 (Shimakami 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

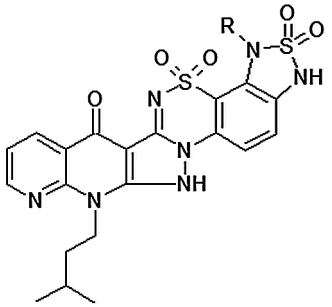
Several research into identifying NS5B inhibitors were carried out and a number of small molecular anti-NS5B compounds have been discovered. These inhibitors are categorized into two general types, namely nucleoside analogs that function as chain terminators, and non-nucleoside inhibitors (NNIs) that 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.

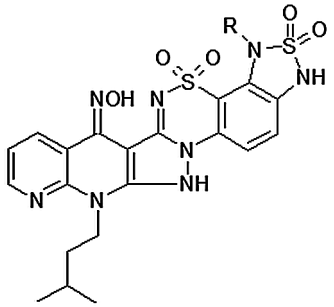
Melagraki *et al.* asserted the results of a ligand-based virtual screening on 98 compounds (benzothiadiazine analogs) which assessed as inhibitors of genotype 1 HCV polymerase (Melagraki et al., 2007). The dataset was divided into a training set of 60 compounds, and a validation set of 38 compounds 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,

EBSCOhost - printed on 2/14/2023 7:16 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

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 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 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

3D QSAR Approach

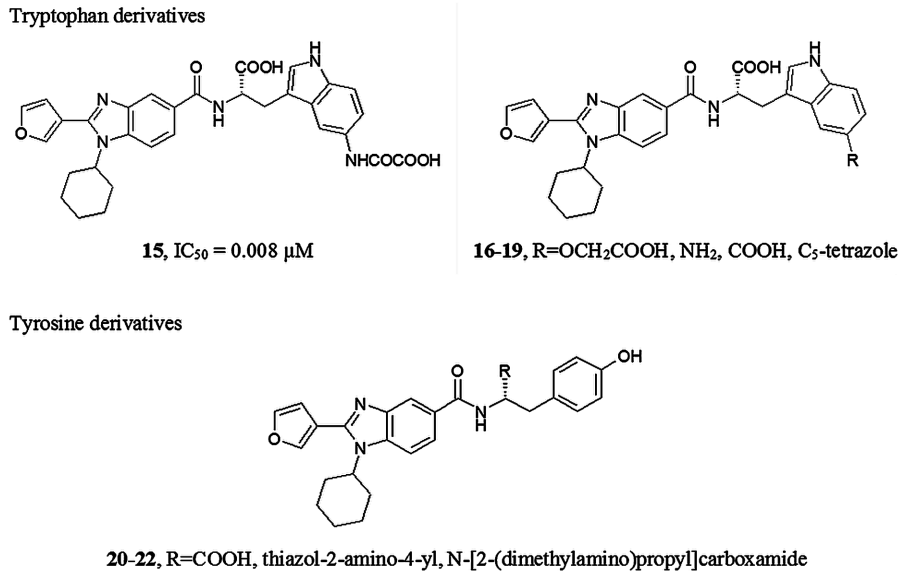
In their recent publication, Patel *et al.* conducted a study of benzimidazole derivatives as hepatitis C virus NS5B polymerase inhibitors using three-dimensional quantitative structure-activity relationship (3D-QSAR) and molecular docking method (Patel et al., 2008). They applied comparative molecular 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.

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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 3. Potent NS5B inhibitors
Presented in (Patel et al., 2008).



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 2/14/2023 7:16 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).

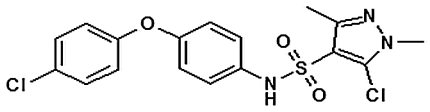
VIRTUAL SCREENING METHOD IN IDENTIFYING NS5B POLYMERASE INHIBITORS

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

Figure 4. Potent NS5B inhibitor
Proposed by (Kim et al., 2008).



Compound **23**

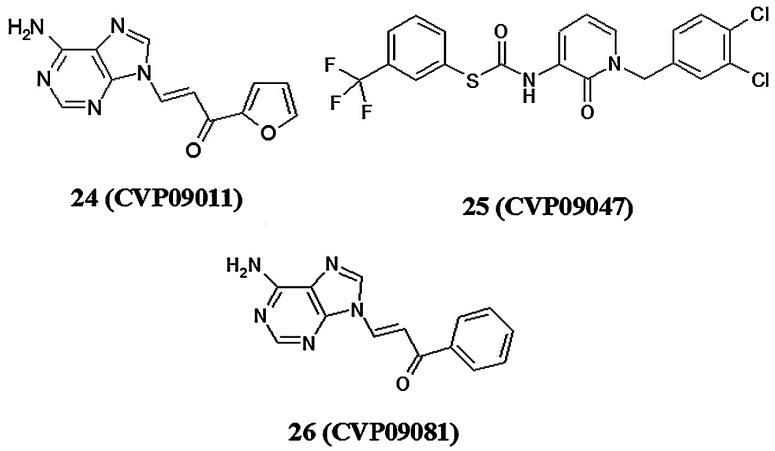
EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 (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

Figure 5. Chemical structure of the novel inhibitors
Proposed by (Ryu et al., 2009).



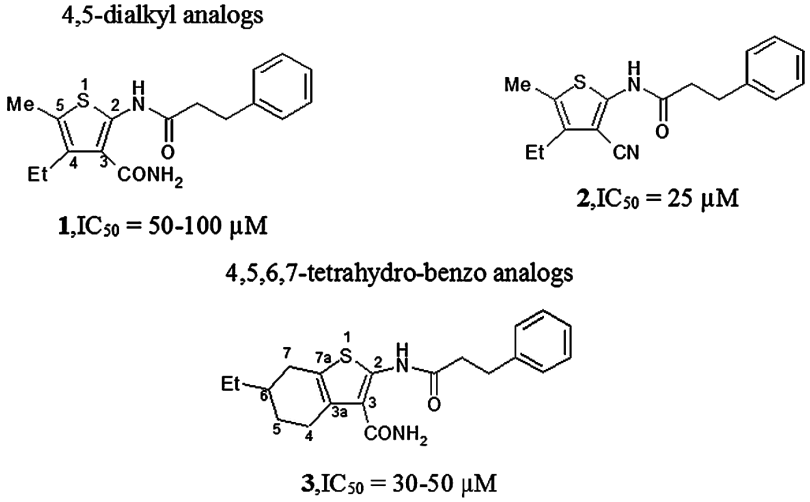
EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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**) 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).

Figure 6. Potent HCV NS5B inhibitors
Proposed from (Louise-May et al., 2007).

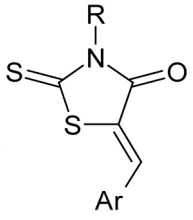


EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 (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).

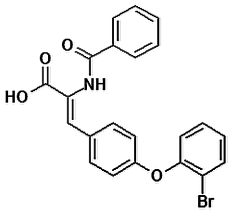
Table 2. Potent NS5B inhibitors proposed by (Talele et al., 2010)



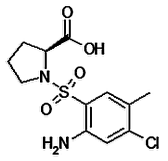
Compound	R	Ar	IC_{50} (μM)
31	$-CH_2COOH$	4-CF_3Ph	55.2 ± 1.10
32	$-CH(COOH)CH_2Ph$	2,4-ClPh	10.6 ± 1.5
33	$-CH_2COOH$	3,4-ClPh	8.4 ± 1.5
34	$-CH_2COOH$	3-PhenoxyPh	7.7 ± 1.4

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

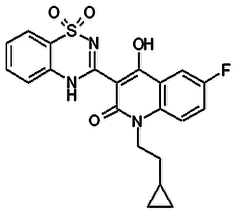
Figure 7. Structures of the five NNIs
Proposed by (T. Li et al., 2010).



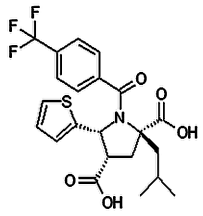
31, 1YVF (acrylic acid derivative)



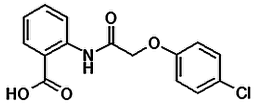
32, 2GC8 (proline sulfonamide derivative)



33, 2GIQ (thiadiazine derivative)



34, 2JC0 (acyl pyrrolidine derivative)



35, 2QE5 (anthranilic acid derivative)

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 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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 $P_{4'}$. 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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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

Figure 8. General structure of peptidomimetic ketoamide derivatives

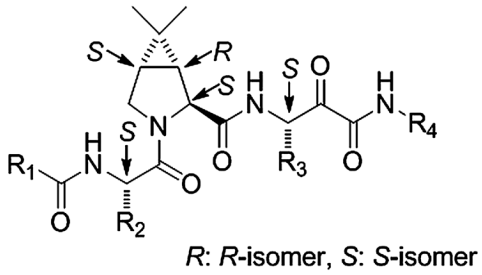
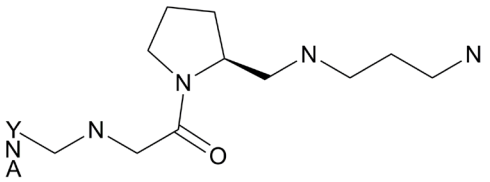


Figure 9. Common substructure used for superimposing the compounds in the data set



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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.

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.

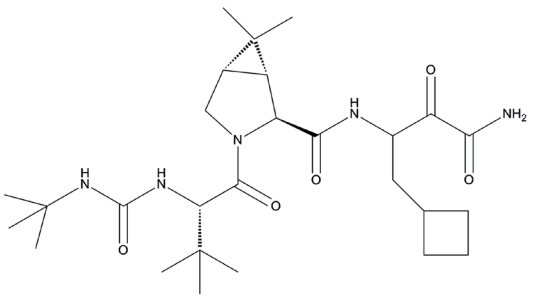
EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 criteria, 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.

Figure 10. Structure of boceprevir and definition of the pharmacophoric groups (P_1, P_2, P_3 and P_4)



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

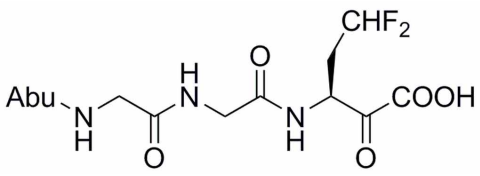
As for the docking process, the most likely conformation of the ligand can be generated. The ligand conformation was identified by estimating the energy of its interaction with the target and the best (*i.e.* the lowest energy) are then returned for further analysis. The docking scoring function, the 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 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.

Figure 11. Compound 2 (abu: 2-aminobutyric acid)



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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.

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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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:

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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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.

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 that 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 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.

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 12. Compound N1

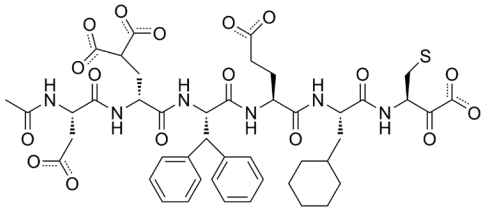


Figure 13. Compound I11

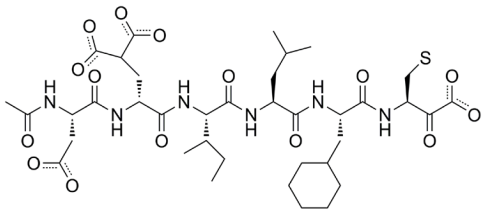
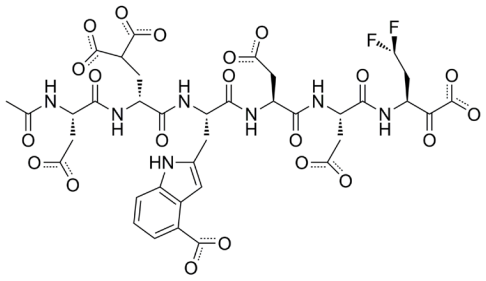


Figure 14. Compound C2



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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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.

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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 10
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: 10.4018/978-1-4666-8136-1.ch010

Copyright © 2015, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/14/2023 7:16 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.

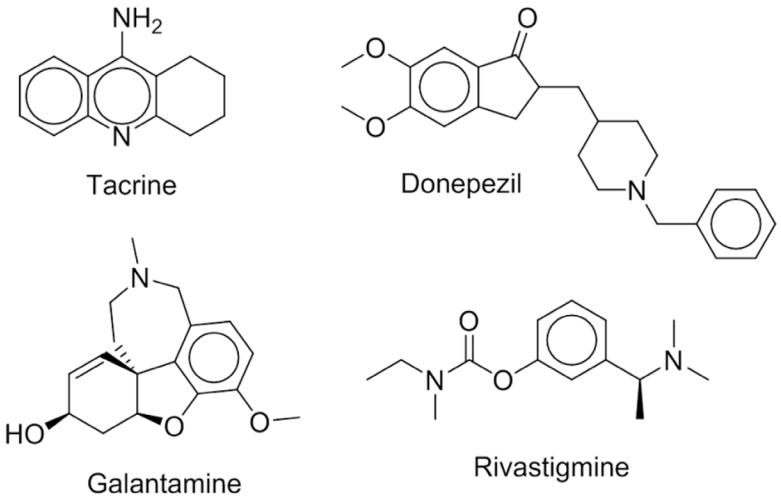
Several evidences showed that reactive oxygen species (ROS) are involved in most of the neurodegenerative pathologies. ROS is the most dangerous reactive species, and which can spoil most of the biological molecules. Hence, it is crucial to maintain the oxidative balance and control in the brain. This 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 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 1. Acetylcholinesterase inhibitors approved by FDA for the treatment of AD



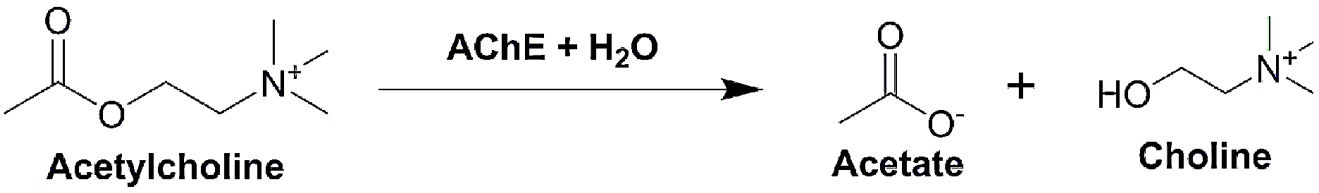
Cholinesterase inhibitors such as tacrine, rivastigmine, donepezil and galantamine, which block the breakdown of ACh has been used for the treatment of mild to moderate AD (Mayeux & Sano, 1999; Silmana & Sussman, 2008, Mirjana, 2013). All of these drugs shown in Figure 1 are reversible AChE inhibitors (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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 2. Enzymatic hydrolysis of ACh by AChE



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.

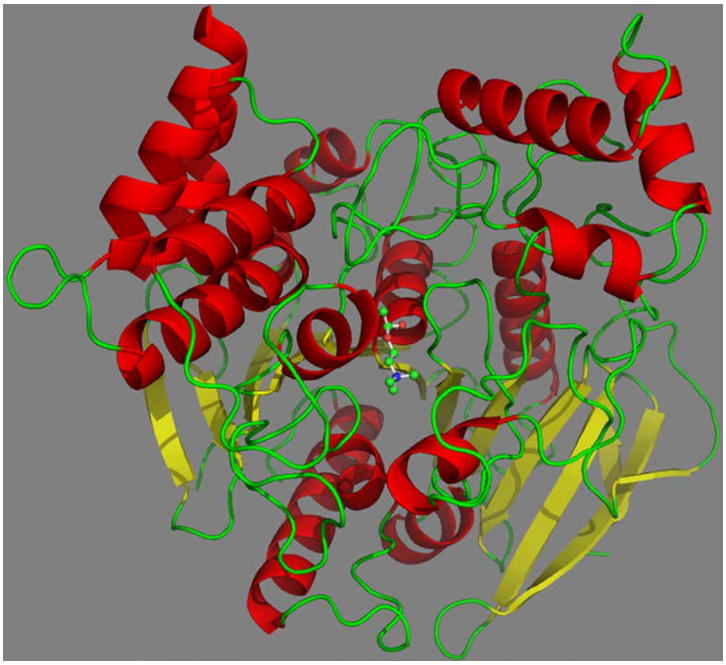
AChE is mainly found at neuromuscular junctions and cholinergic synapses in the central nervous system, where its activity serves to terminate the synaptic transmission. AChE has a very high catalytic activity, and each molecule of AChE degrades about 5000 molecules of ACh per second. The choline produced 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.

EBSCOhost - printed on 2/14/2023 7:16 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.



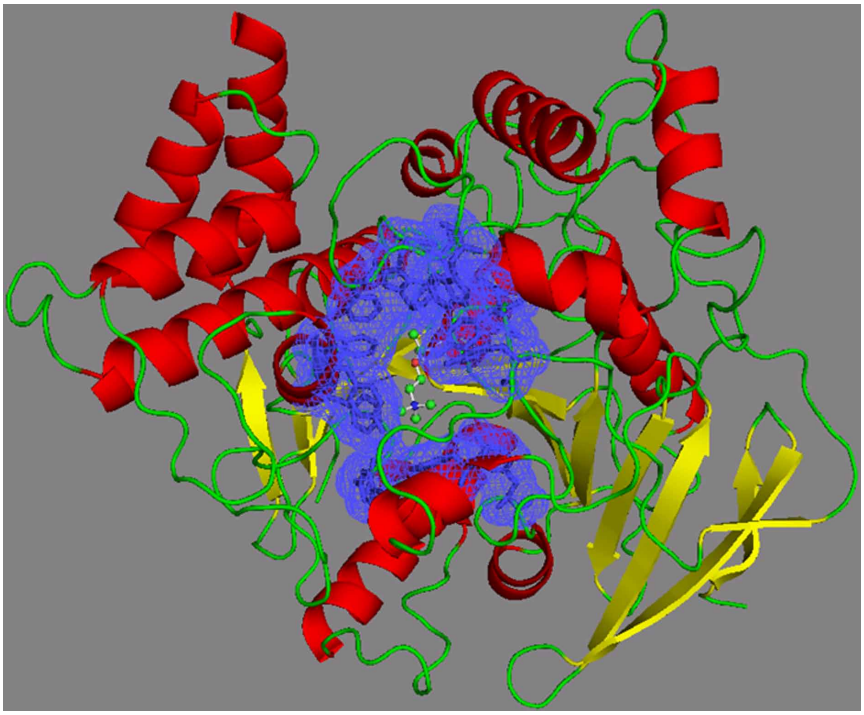
The Active-Site Gorge and the Catalytic Triad

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 conserved 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

EBSCOhost - printed on 2/14/2023 7:16 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.



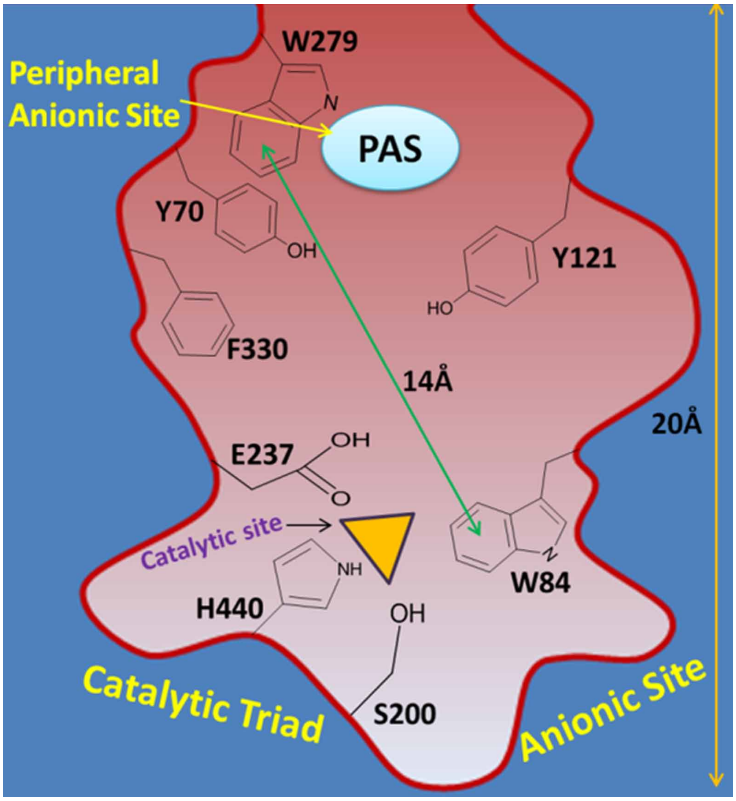
residues, and their flanking sequences, which are highly conserved in AChEIs from different species was N66 and I444. They are synthesized on the first exon, which codes for residues 1–480. It should be noted that the gorge contains only a few acidic residues, which include D285 and E273 at the very 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, Ekström, & 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

EBSCOhost - printed on 2/14/2023 7:16 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).



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 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, &

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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

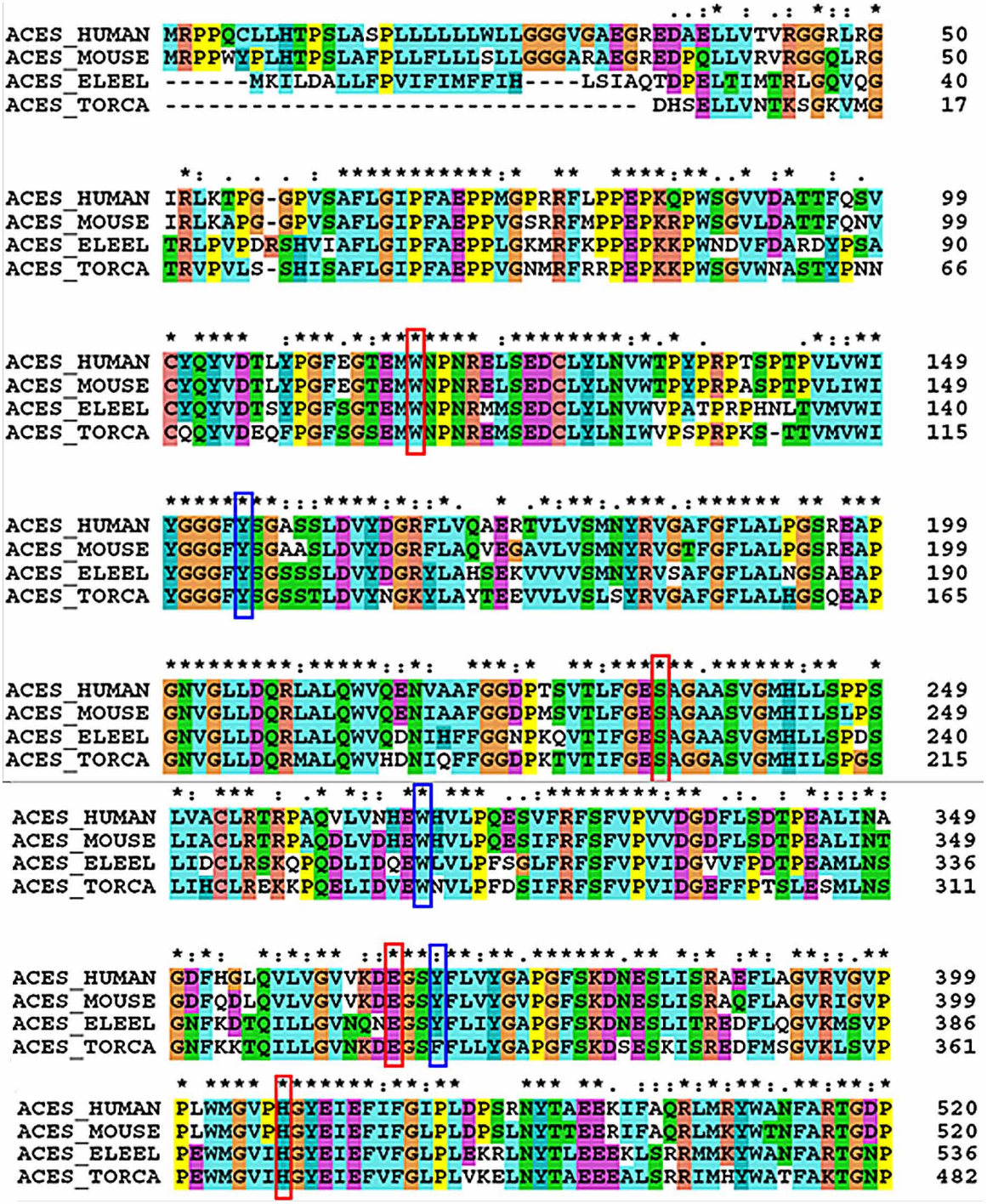
AChEIs can be classified into different groups, based on their mode of interaction with the AChE enzyme. Inhibition can be either reversible, by competitively preventing the substrate from reaching the active site; pseudo-irreversible, by covalent reaction with the active site serine, inactivating the 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 2/14/2023 7:16 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 2/14/2023 7:16 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.

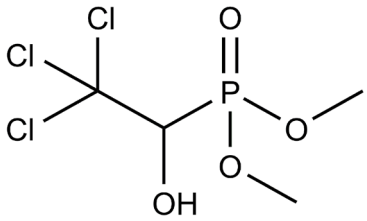
Pseudo-Irreversible AChEIs

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 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 8. Chemical structure of pseudo-irreversible inhibitors of AChE

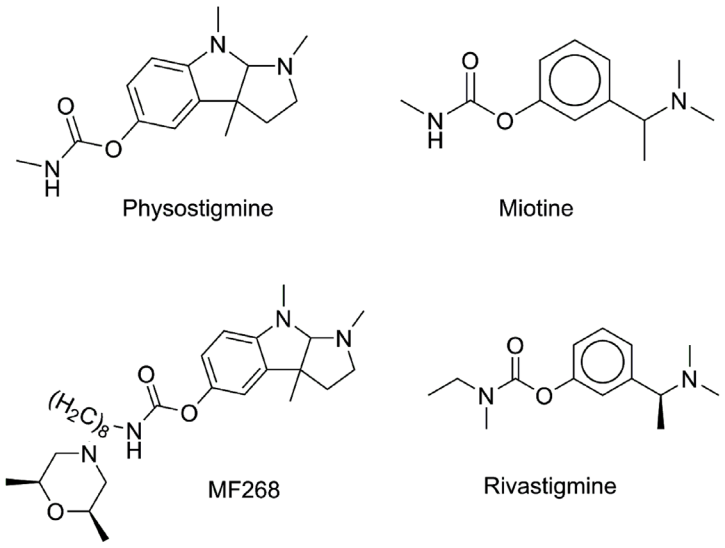
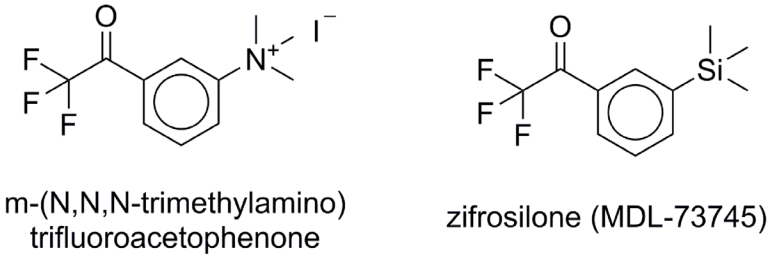


Figure 9. Chemical structure of m-(N,N,N-trimethylamino) trifluoroacetophenone and zifrosilone (MDL-73745) as transition state analogue of AChEIs



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 moderately 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).

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 10. Chemical structure of tacrine, velnacrine, and suronacrine

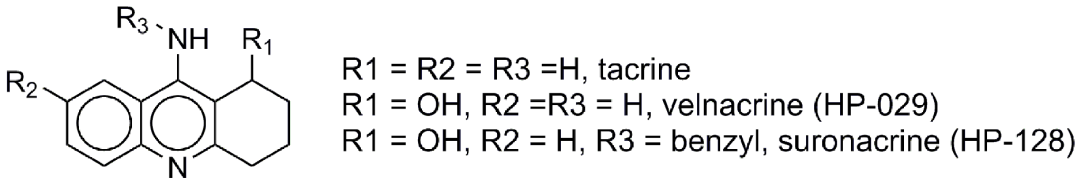
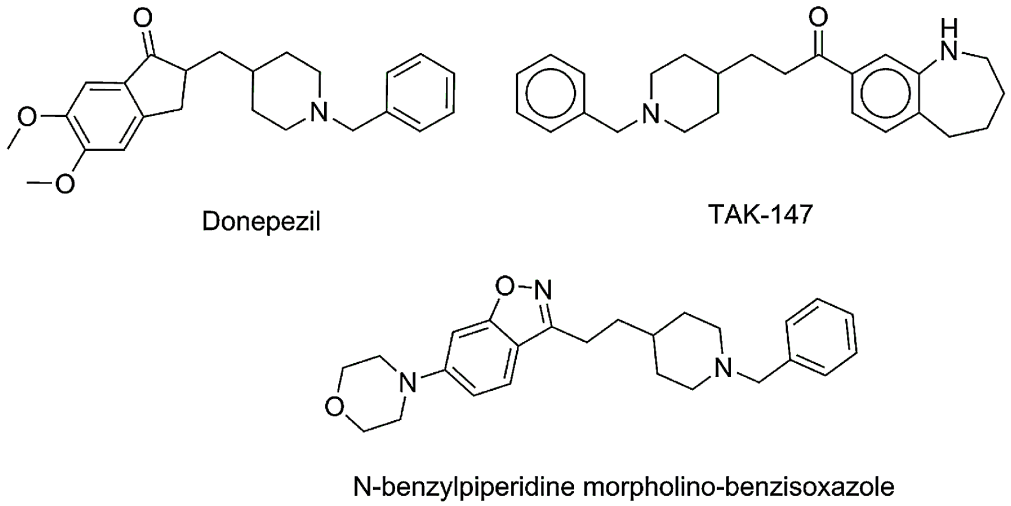


Figure 11. Structure of donepezil (E2020), TAK-147 and N- benzylpiperidine morpholino-benzisoxazole



- **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).

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

- **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).

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

Figure 12. Chemical structure of galanthamine, P11012 and P11149

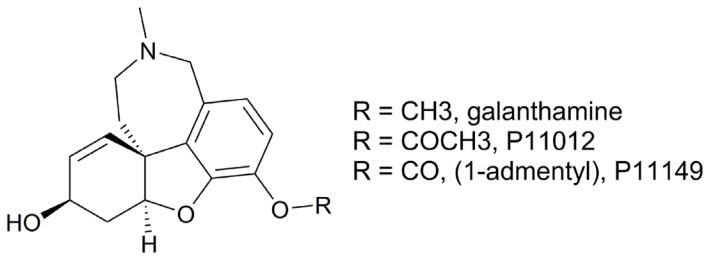
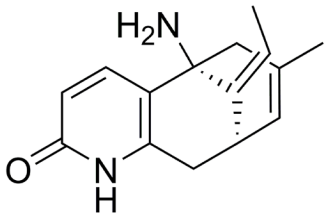
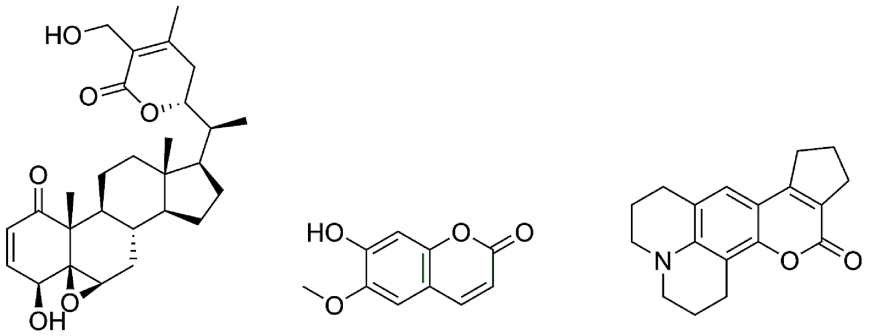


Figure 13. Chemical structure of (-) -huperzine a



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 14. Chemical structure of withaferin a, scopoletin, and coumarin106



Withaferin A Scopoletin Coumarin106

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 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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

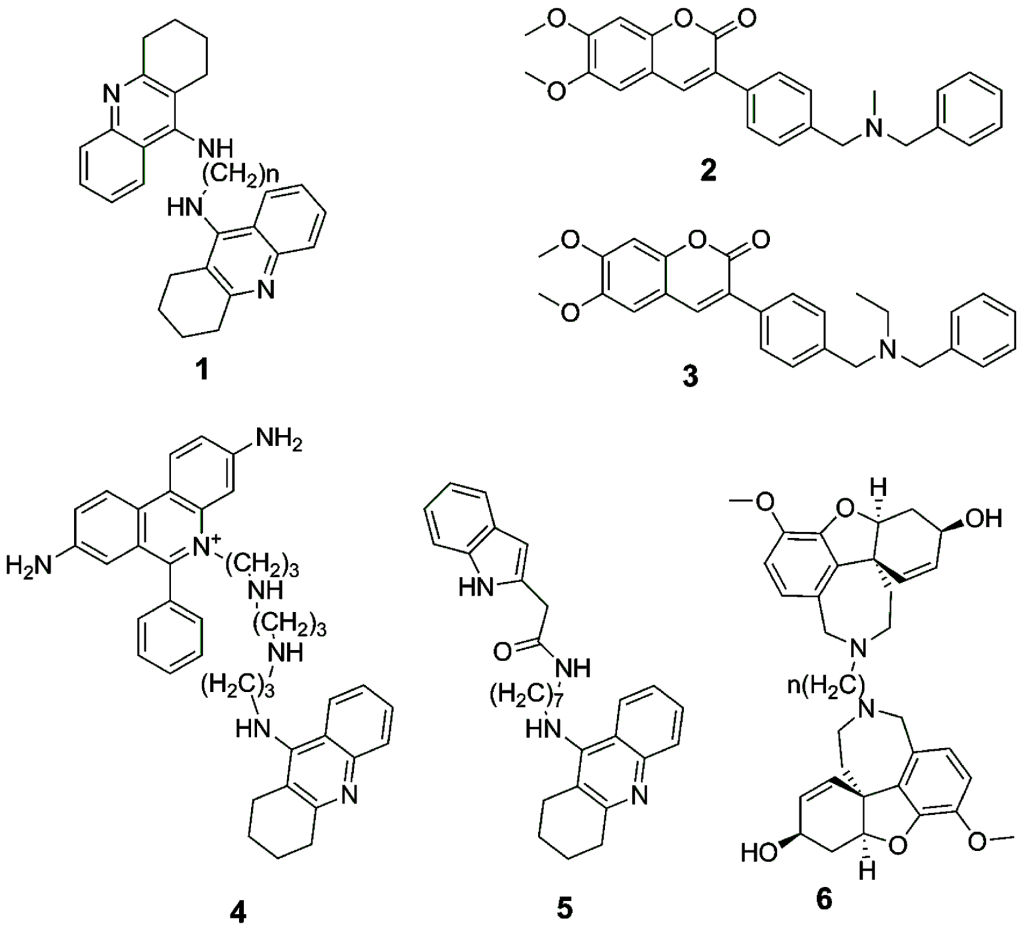
100 μM it can inhibit 22% of the AChE-induced aggregation of Aβ (Bartolini, Bertucci, Cavrini, & Andrisano, 2003). In 1996, the first family of a dual binding site AChEIs was developed by the group of Pang and Carlier, in which the two component units are connected through a heptamethylene chain (Pang, 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).

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).

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 15. Different dual binding site AChE inhibitors with Aβ anti-aggregating effects



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:

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:

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 gave rise to the "$\sigma-\rho$" 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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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 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,

$$LOF = LSE \,/\, \left\{1 - \left(c + dp\right) / \, m\right\}^2 \tag{1}$$

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

$$y_i = w_i x_i + b \tag{2}$$

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:

$$Y = f\left(x\right) = \sum\nolimits_{i=1}^{n} w_i \dot{\phi}\left(x_i\right) + b \tag{3}$$

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:

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

$$R(C) = C\frac{1}{N}\sum_{i=1}^{N} L\varepsilon(y_i, f(x_i, w)) + \frac{1}{2}\left\|\mathrm{w}\right\|^2 \quad (4)$$

$$L\varepsilon\left(y, f\left(x, w\right)\right) = \begin{matrix} \left\{\left|y - f\left(x,w\right)\right|\right\} - \varepsilon, & for\ \left|y - f\left(x,w\right)\right| \geq \varepsilon \\ 0 & otherwise \end{matrix}$$

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.

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:

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y}_{Training})^2} \tag{5}$$

$$R_a^2 = \frac{(N-1)R^2 - d - 1}{N - d - 1} \tag{6}$$

$$F = \frac{(R^2 \,/\, d)}{((1-R^2)\,/\,(N-d-1))}. \tag{7}$$

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 $\mathrm{R}^2_{\mathrm{pred}}$, 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:

$$R_{pred}^2 = 1 - \frac{\sum\left(\mathrm{Y}_{pred(\mathrm{Test})} - Y_{(Test)}\right)^2}{\sum\left(Y_{(Test)} - \bar{Y}_{(Training)}\right)^2}. \tag{8}$$

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_{pred}^2 and r^2 might not be sufficient to indicate the external 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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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^2_{LOO}, 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^2_{LOO}, and S_{PRESS}, is calculated according to the formula defined below:

$$R^2_{LOO} = 1 - \frac{\sum_{i=1}^{n}(Y_i - \tilde{Y}_{i\,LOO})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y}_{Training})^2} \tag{9}$$

$$S_{PRESS} = \sqrt{\frac{PRESS}{(N-d-1)}} \tag{10}$$

$$RMSE = \sqrt{\frac{PRESS}{N}}\,. \tag{11}$$

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^2_{LOO} value ($R^2_{LOO} > 0.5$) 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. 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 ${}^{c}R^2_p$ metric proposed by Mitra, Saha, and Roy (2010) ${}^{c}R^2_p = R\sqrt{\left(R^2 - R^2_r\right)}$. 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 proper-

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

ties (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 pharmaceutical 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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

(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.

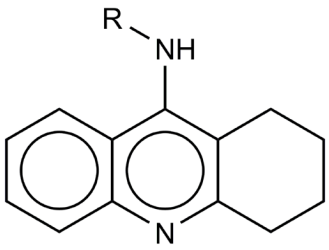
Tacrine (9-amino-1,2,3,4-tetrahydroacridine, THA) as the first anti-Alzheimer's drug approved by the FDA in 1993, which has beneficial effects on cognition in patients with AD (Knapp, Knopman, Solomon, Pendlebury, Davis, & Gracon, 1994). Tacrine was described as a reversible inhibitor of AChE and BuChE 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 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,$CH{=}CH_2$]

$$\text{Log } 1/Ki = -0.296(0.136)\,L + 0.569(0.220)\log(\beta \times 10^{L} + 1) + 7.921(0.539) \quad (1)$$

$n = 11$, $r^2 = 0.844$, $q^2 = 0.842$, $s = 0.142$, $F_{3,7} = 12.59$, $L_{opt} = 4.903$, $\log \beta = -4.867$

Figure 16. N-monosubstituted 1,2,3,4-tetrahydro-9- aminoacridines (Recanatini, et al., 1997).



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

$$\text{Log } 1/IC_{50} = -0.560(0.132)\,L + 0.877(0.263)\log(\beta \times 10^{L} + 1 + 1.509(0.386)\text{I-Cl} + 7.141(0.528) \quad (2)$$

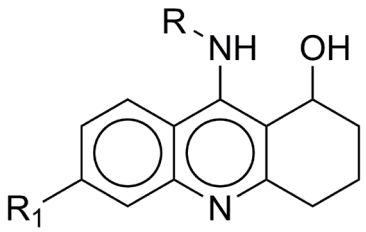
$n = 30$, $r^2 = 0.861$, $q^2 = 0.857$, $s = 0.305$, $F_{4, 25} = 38.83$, $L_{opt} = 5.116$, $\log \beta = -4.869$.

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.

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.

Figure 17. 9-amino-1,2,3,4-tetrahydroacridin-1-ol analogues (Shutske et al. 1989).



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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).

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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).

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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

$$pIC_{50} = 5.793 + 0.443Atype_C_3 + 0.301Atype_C_32 - 0.965Atype_N_70$$

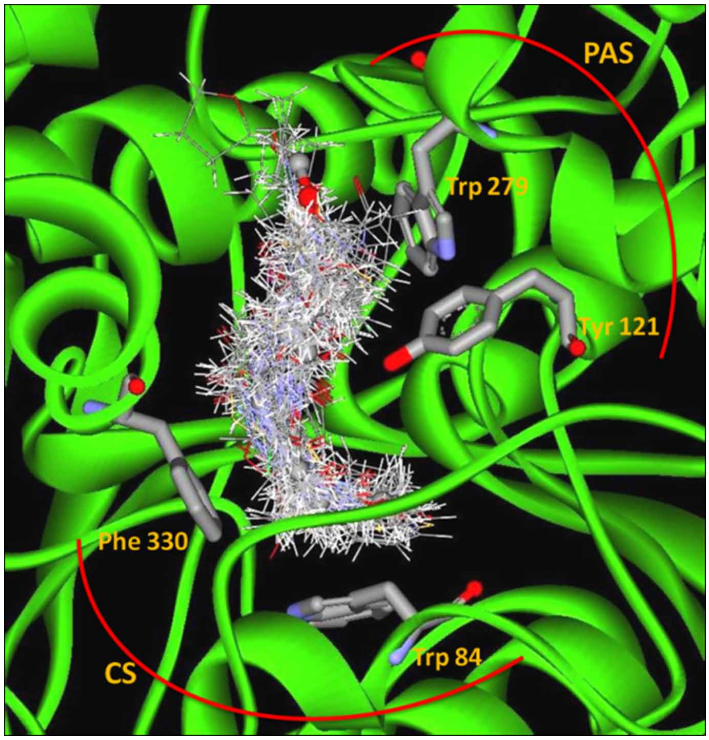
$$- 1.057S_ssS + 0.071Atype_H_46 - 0.215LUMO \quad (6)$$

$N = 31$, $R^2 = 0.887$, LOF = 0.079, $R^2_a = 0.859$, $q^2 = 0.838$, F = 31.44,

$n_{Test} = 11$, $R^2_{pred} = 0.756$, $r^2 = 0.788$, $R^2_0 = 0.705$, $R^2_m = 0.561$.

EBSCOhost - printed on 2/14/2023 7:16 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).



G/PLS Model

$$pIC_{50} = 5.808 + 0.469Atype_C_3 + 0.200Atype_N_75 - 1.446Atype_C_30$$

$$- 1.059S_ssS + 0.077Atype_H_46 - 0.209LUMO \quad (7)$$

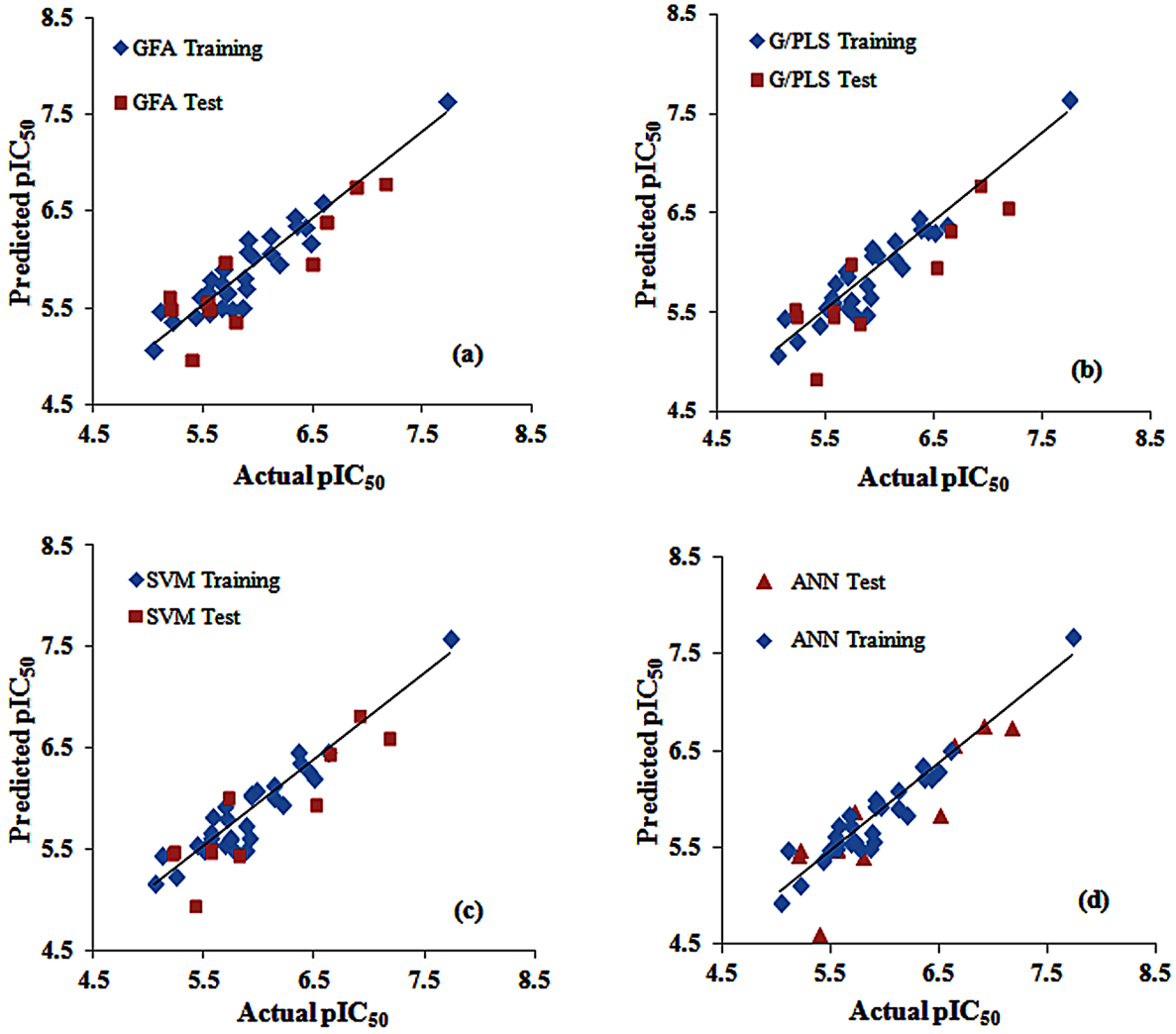
$$N = 31,\ R^2 = 0.889,\ LSE = 0.029,\ R_a^2 = 0.827,\ q^2 = 0.739,\ F = 40.04,$$

$$n_{Test} = 11,\ R_{pred}^2 = 0.706,\ r^2 = 0.757,\ R_0^2 = 0.659,\ R_m^2 = 0.520.$$

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 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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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).



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 (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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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).

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 multi-potent 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).

EBSCOhost - printed on 2/14/2023 7:16 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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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.

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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.

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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.

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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 11
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: 10.4018/978-1-4666-8136-1.ch011

Copyright © 2015, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/14/2023 7:16 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).

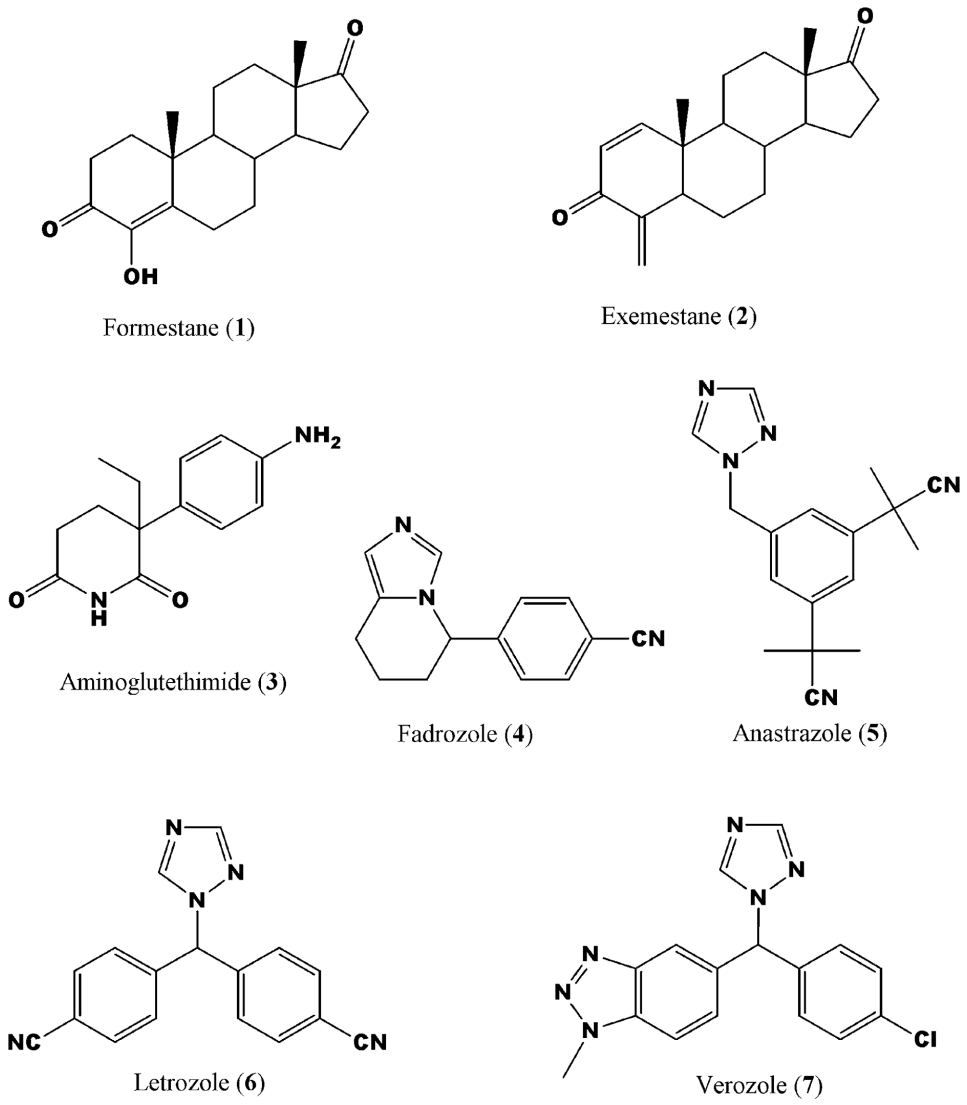
Sex hormones play crucial roles like growth regulation, maturation and reproduction in living animals. Sex hormones are composed of a steroidal cyclopentanoperhydrophenanthrene nucleus. Steroidal hormones are involved in regulation of different physiological effects, like muscle and hair growth, fertility, 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;

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Osborne & Schiff, 2005; Recanatini & Cavalli, 1998; Recanatini, 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

Figure 1. Aromatase inhibitors (AIs) evaluated in clinical trials



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

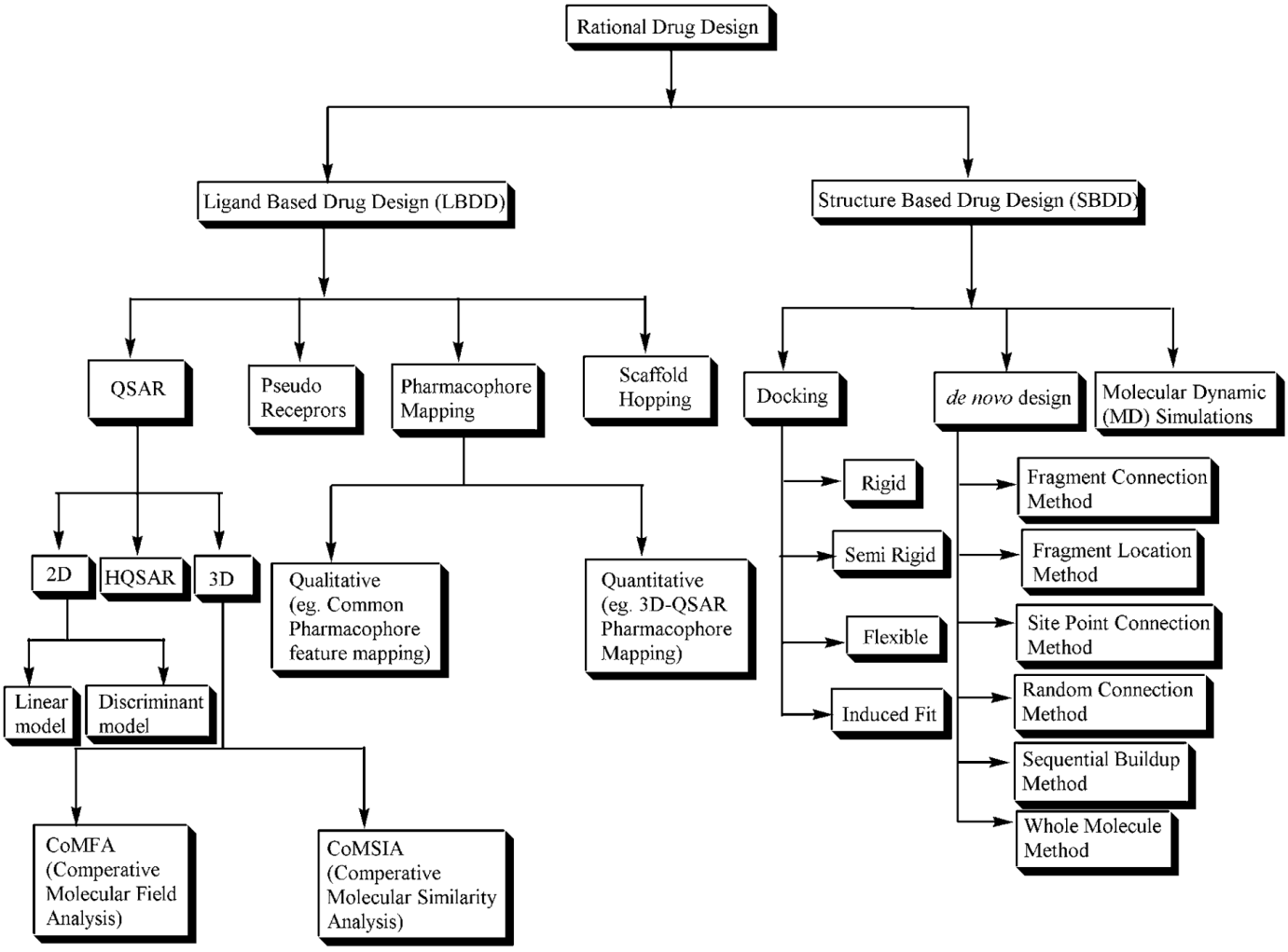
hand, nonsteroidal aromatase inhibitors (NSAIs) bind noncovalently to the aromatase enzyme. These block the active site and reversibly inhibit the enzyme (Brueggemeier, Hackett, & Diaz-Cruz, 2005; Hong *et al*., 2007). The NSAIs are divided according to the order of their discovery. Currently, 3rd generation 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 2/14/2023 7:16 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,

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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.

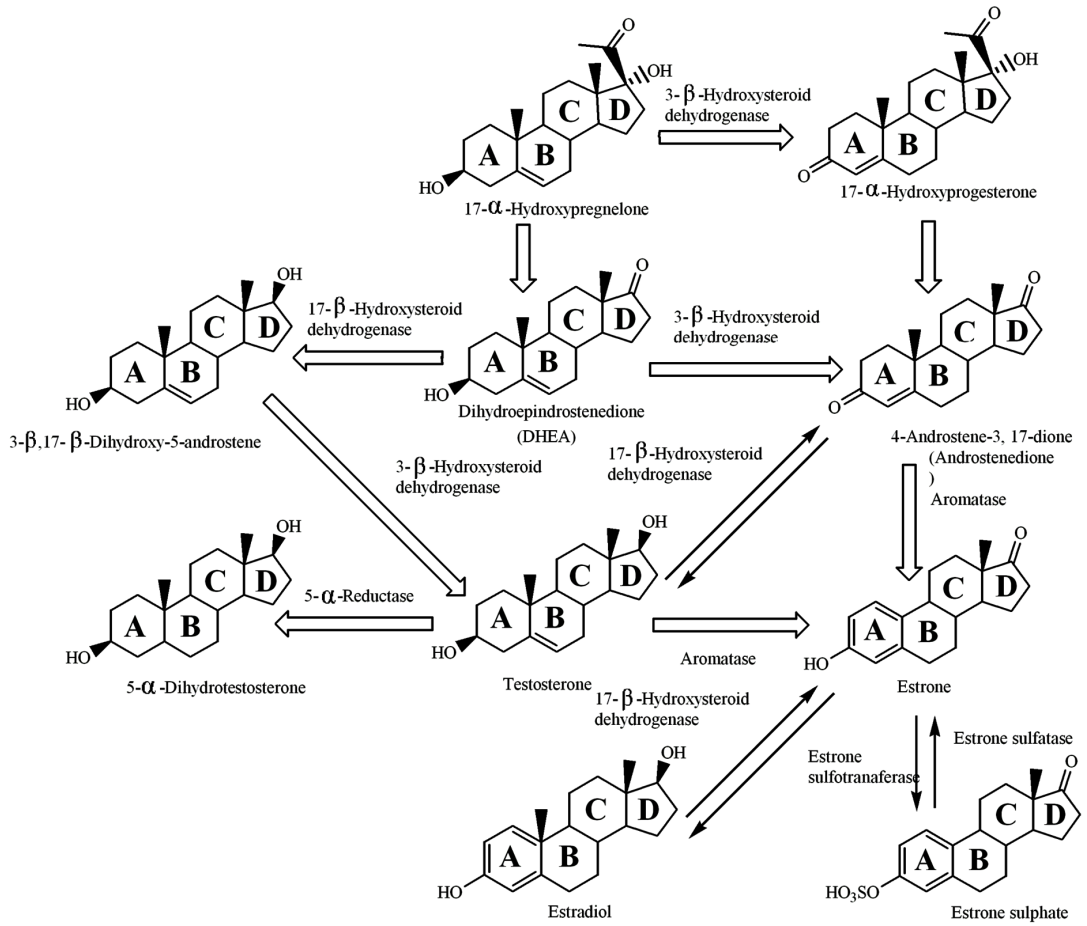
AROMATASE BIOCHEMISTRY

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 &

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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 biosynthesis of estrogens from androgens (Figure 3) (Schuster *et al*., 2006).

Figure 3. Biosynthetic pathway of estrogens and androgens



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

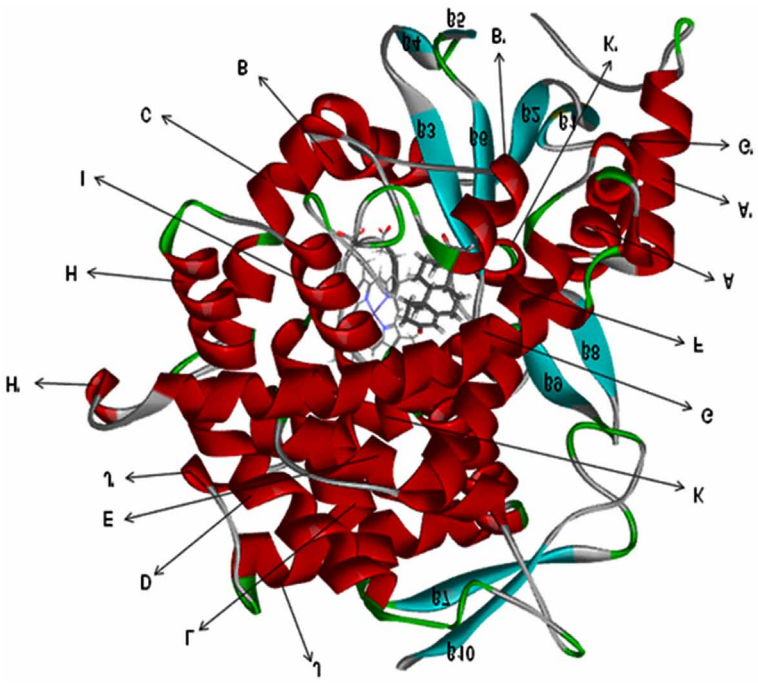
Estradiol is the most important endogenous estrogen which is biosynthesized from androgens by the cytochrome P450 enzyme complex known as 'aromatase' (Simpson *et al*., 1994). Aromatase is a multienzymatic complex which is mostly expressed in the ovary of premenopausal women, placenta, peripheral adipose 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.

EBSCOhost - printed on 2/14/2023 7:16 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).



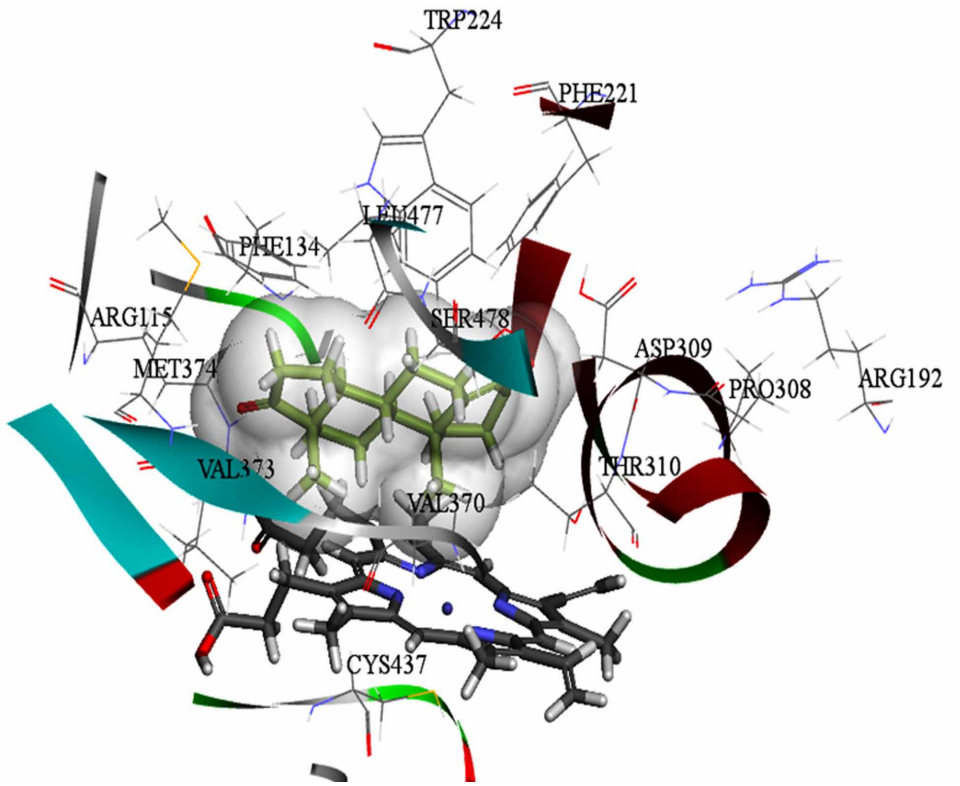
The tertiary structure of human aromatase is made up of 12 major α-helices as well as 10 β-strands. These β-strands are distributed into 1 major and 3 minor sheets. These show the characteristic cytochrome P450 fold (Ghosh *et al*., 2010). The amino terminus starts at residue 45, whereas the carboxyl 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 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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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).

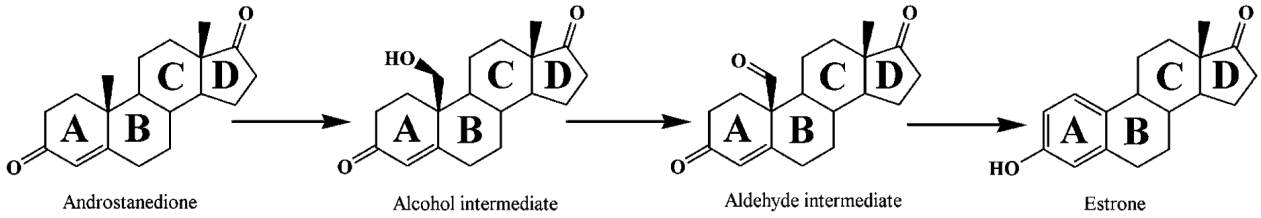
The refined Fe-position is displaced around 0.2 Å away from the heme plane towards the amino acid residue Cys437 of aromatase enzyme. The oxygen atom of 17-keto group of the substrate is located at a distance 2.8 Å from the backbone amide nitrogen of Met374. This helps to accept a proton as well as makes a hydrogen bond. The 3-keto oxygen on the other end is at 2.6 Å away from the carboxylate oxy-

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 2/14/2023 7:16 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).



gen 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 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.

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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 7. Possible mechanism for H2β abstraction and 2,3-enolization of androstenedione (Ghosh et al., 2009).

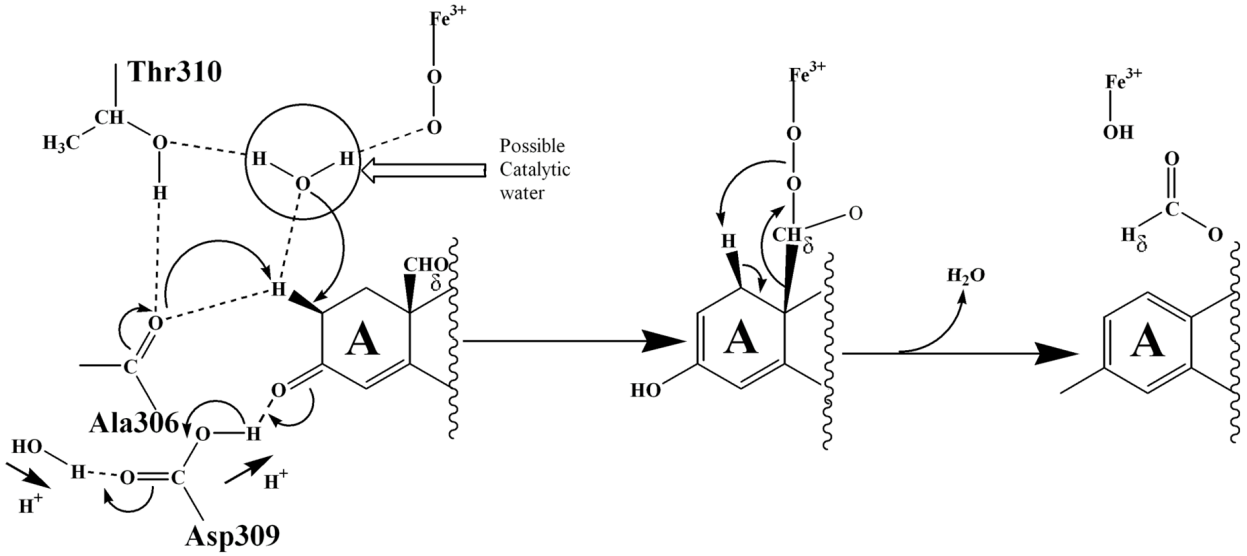
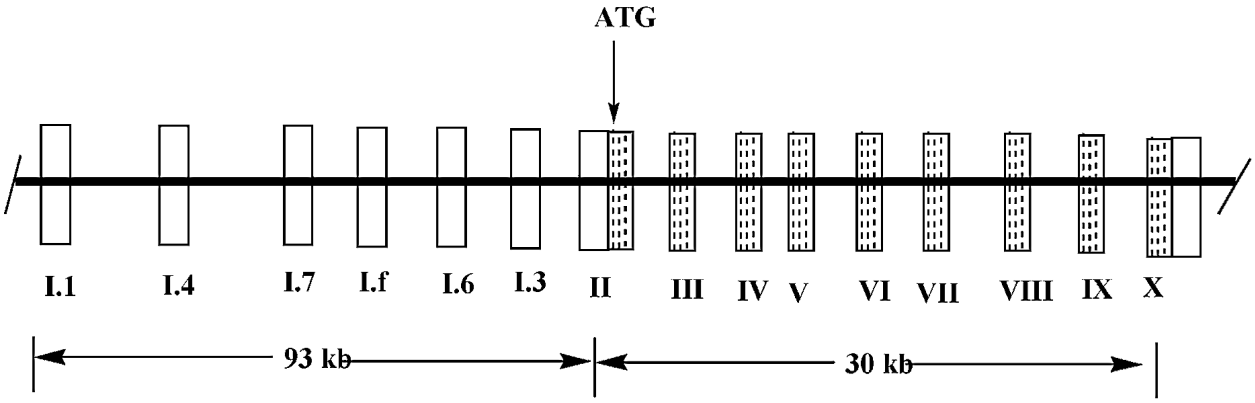


Figure 8. Tissue specific aromatase gene and promoter regions



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

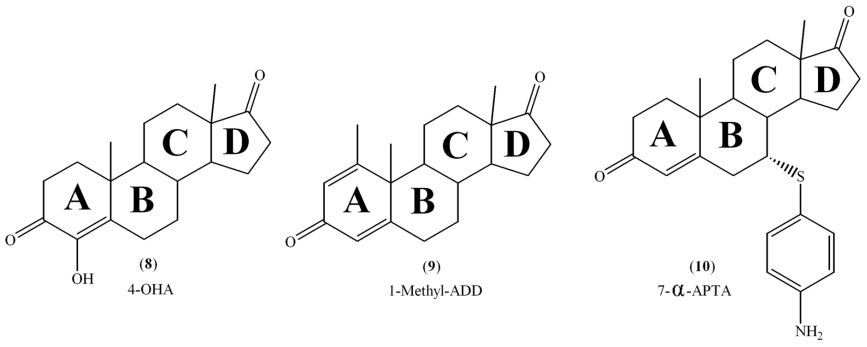
EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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.

AROMATASE INHIBITORS (AIs)

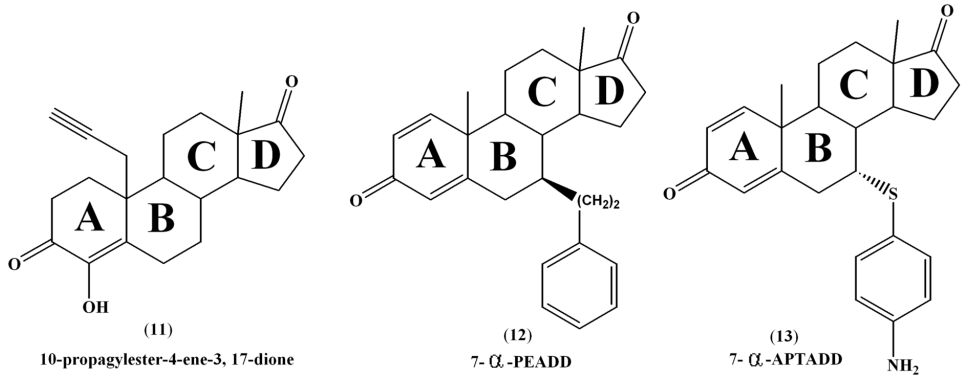
Aromatase inhibitors are mainly categorized into three generations based on their chronology of clinical trials. These are also classified as type-I and type-II inhibitors depending on their mechanisms of action. Aminoglutethimide (AGT) (**3**) (Figure 1) is the first generation AI, whereas the second 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).

Figure 9. Steroidal competitive enzyme inhibitors



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 10. Steroidal mechanism-based aromatase inhibitors

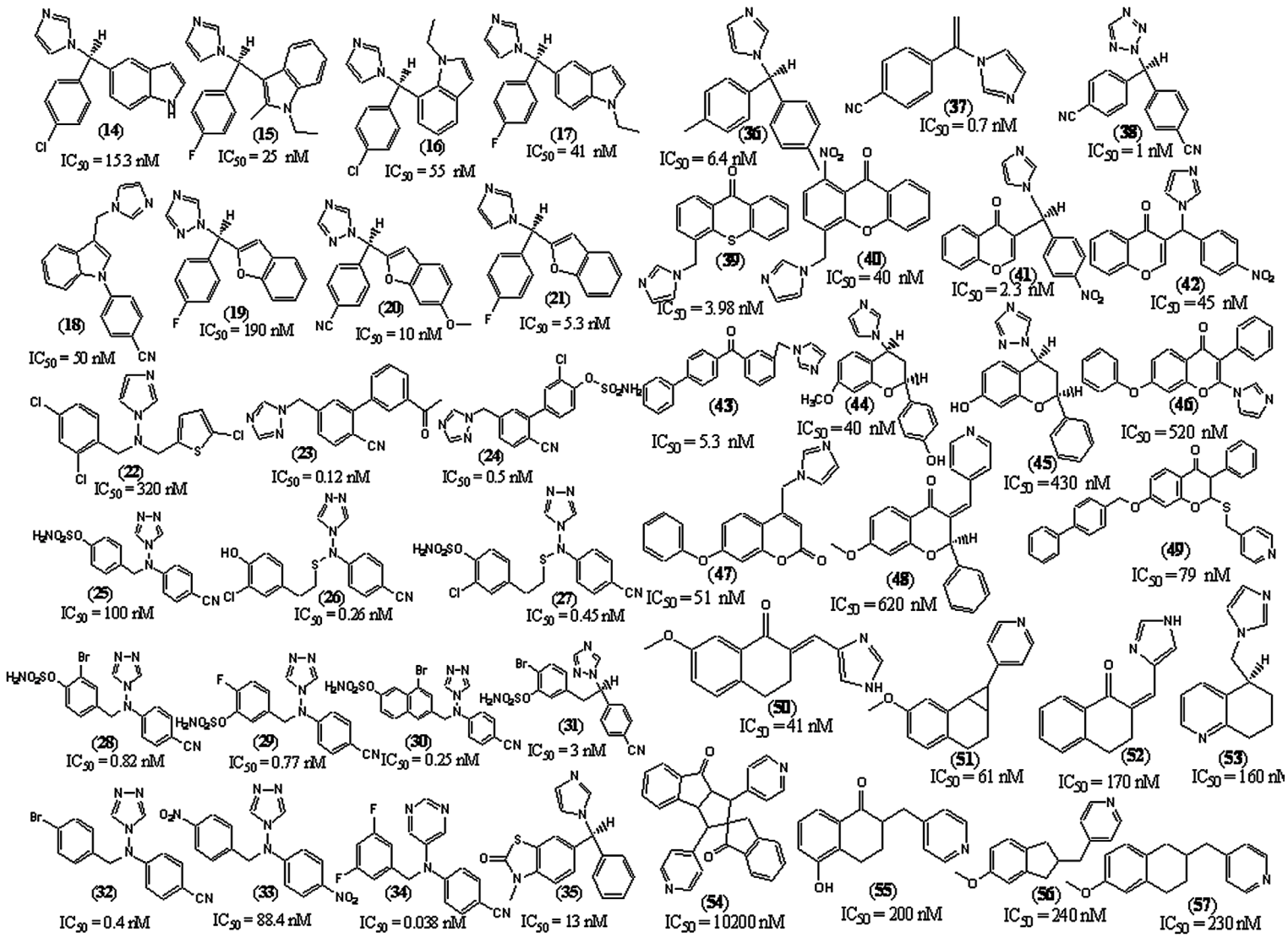


The mechanism-based inhibitors mimic the substrate and are transformed by the enzyme to reactive intermediates. These result in inactivation of aromatase enzyme. These inhibitors, in other terms are called as 'enzyme-activated irreversible inhibitors' or 'suicide substrates' or 'suicide inactivators'. 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*., 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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 11. Selective potent non-steroidal aromatase inhibitors (NSAIs)



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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 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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 12. Structural requirements of chromones for effective NSAIs

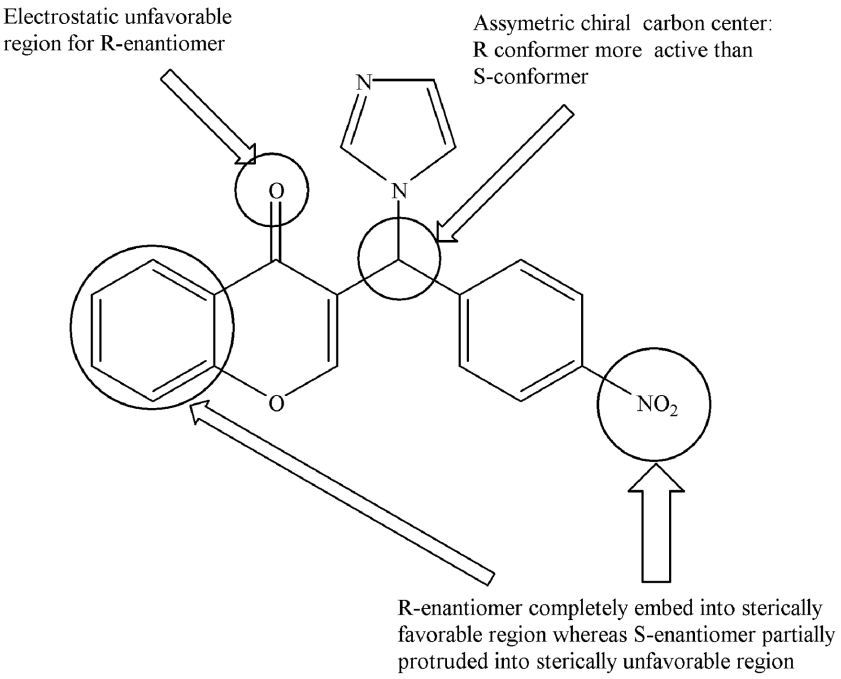
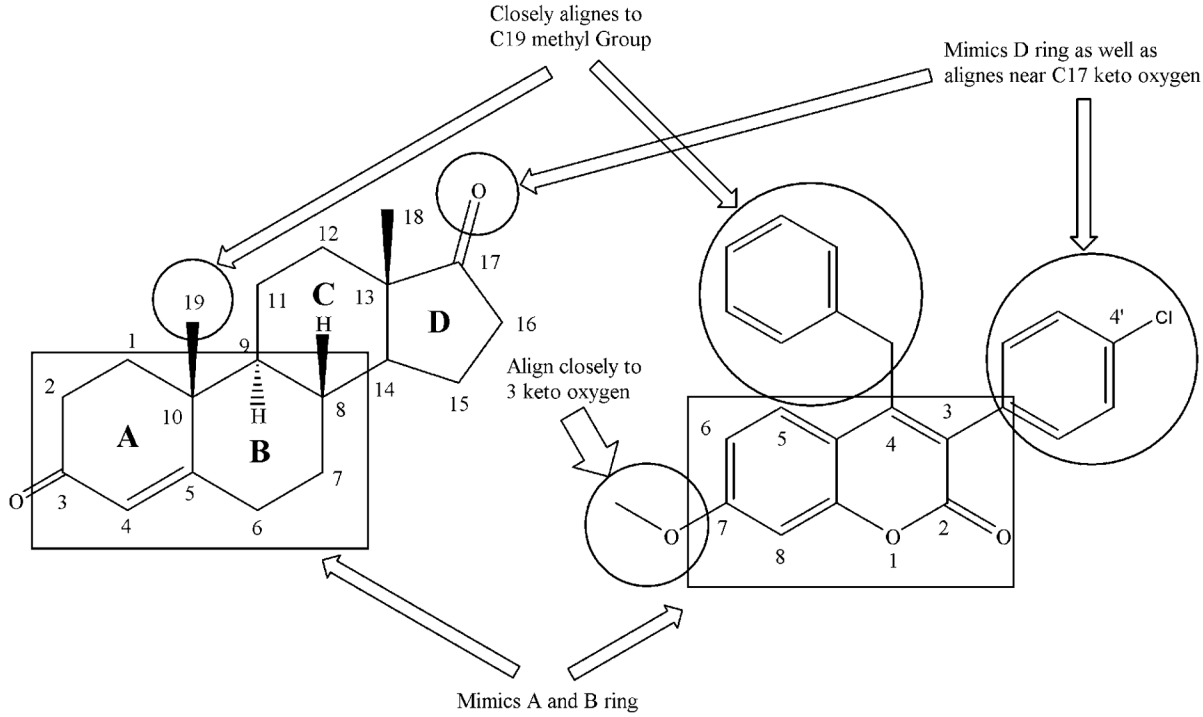


Figure 13. Crucial structural requirements of coumarins as promising NSAIs

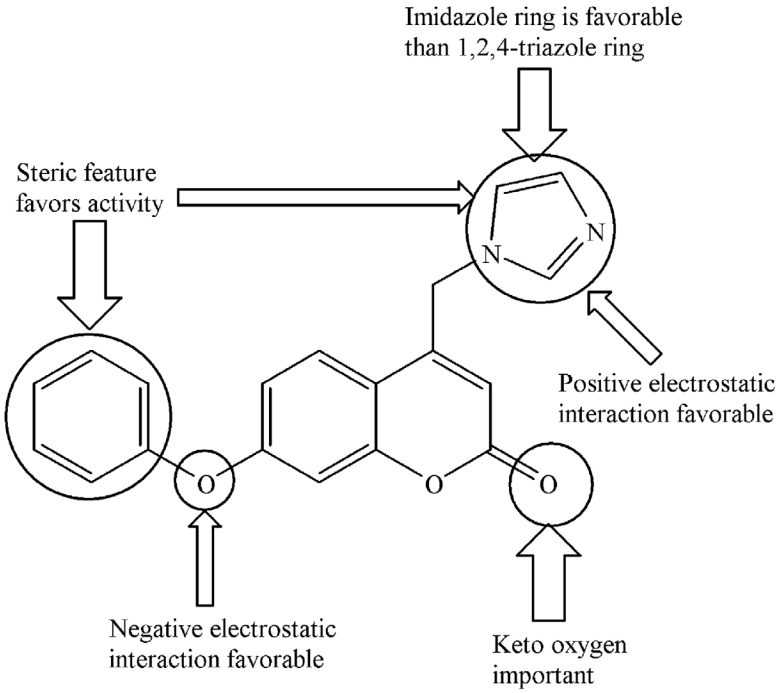


EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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

Figure 14. Important structural features of coumarin azoles required for better NSAIs

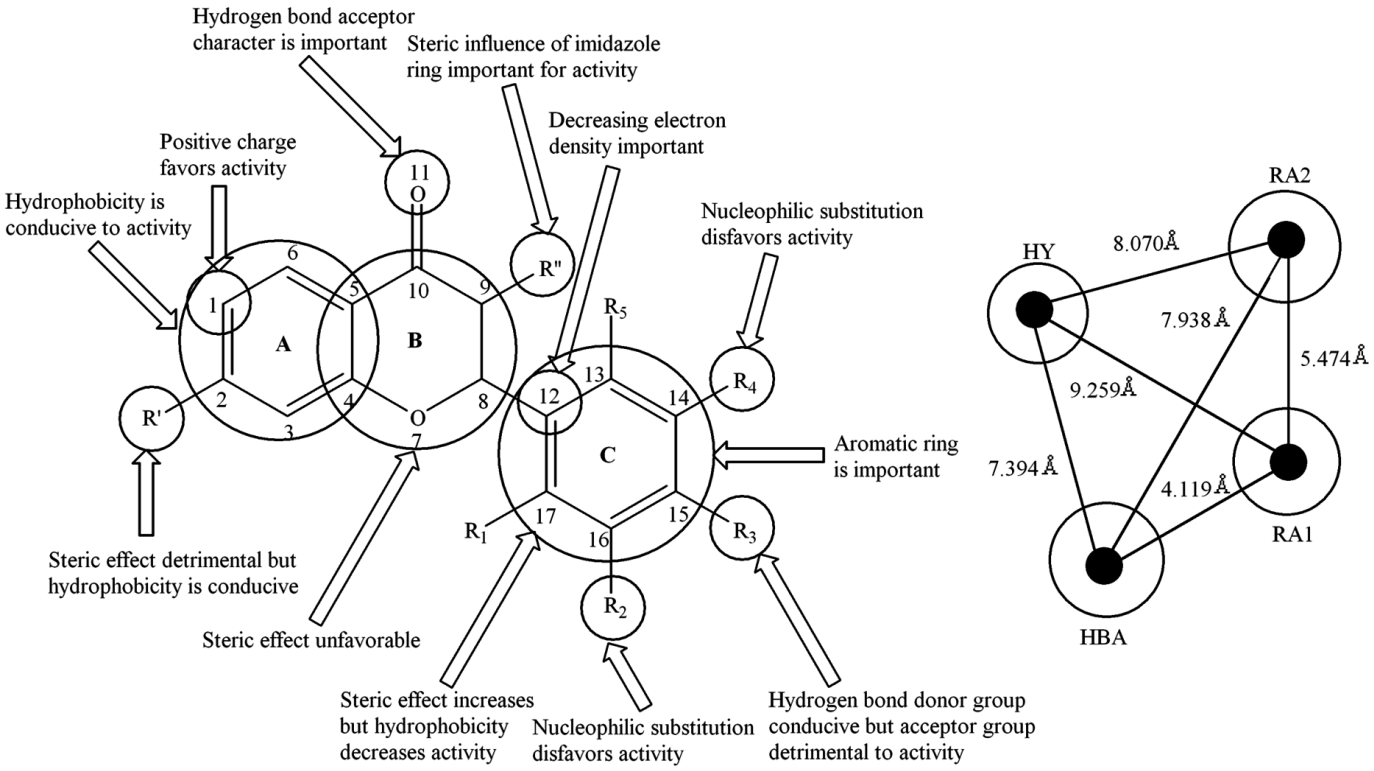


EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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,

Figure 15. Structural and pharmacophoric requirements of flavones for aromatase inhibitory activity



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

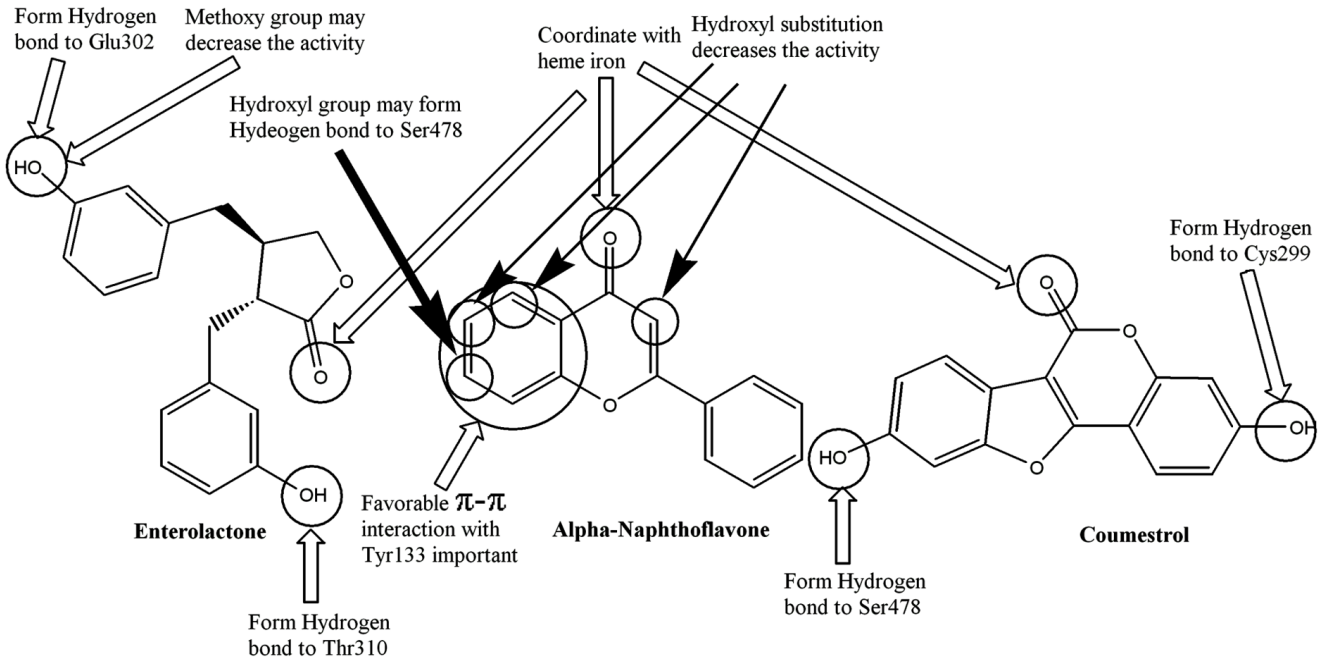
2009). The simulation study was analyzed to find the essential interactions took place on binding with the aromatase enzyme to the phytoestrogen, and were subjected to docking study to know the binding interactions. The MDS study of the aromatase-enterolactone complex suggested that there was only an 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 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, 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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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.

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 stabil-

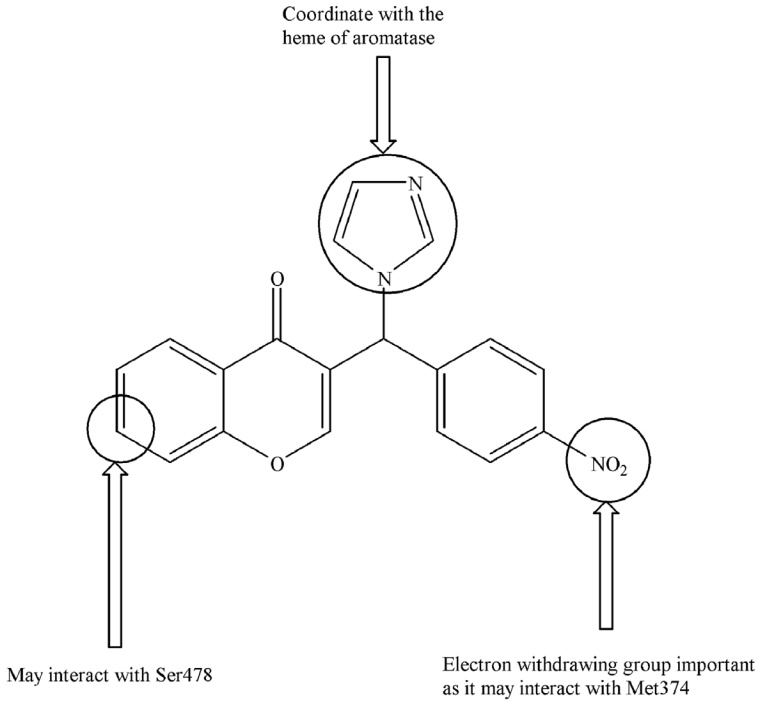
Figure 16. Schematic representation of the docking results of phytoestrogens for better activity



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

ity 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 pair-wise 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.

Figure 17. Important structural features of flavonoids for potent anti-aromatase activity

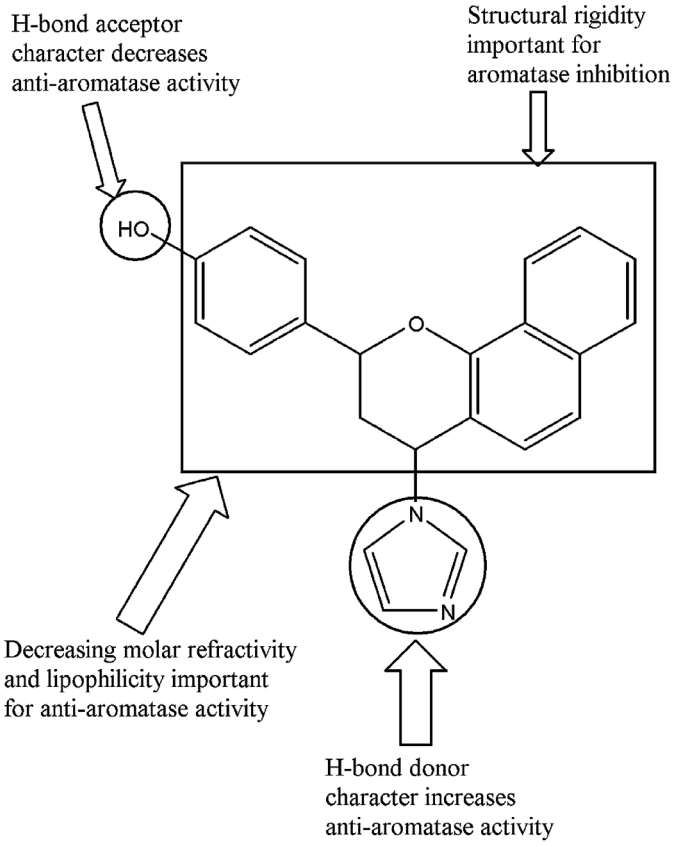


EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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

Figure 18. Structural requirements of flavonoids for potential anti-aromatase activity

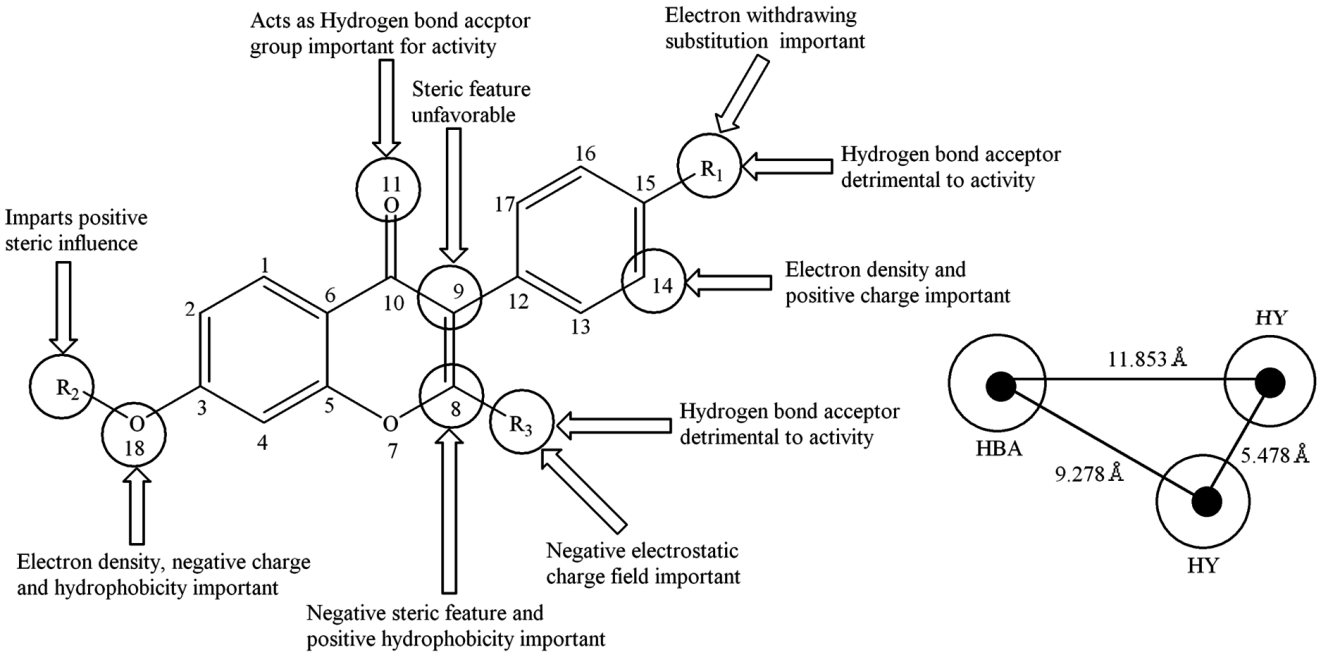


EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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.

Figure 19. Structural and pharmacophoric requirements of isoflavones for anti-aromatase activity



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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.

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.

EBSCOhost - printed on 2/14/2023 7:16 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

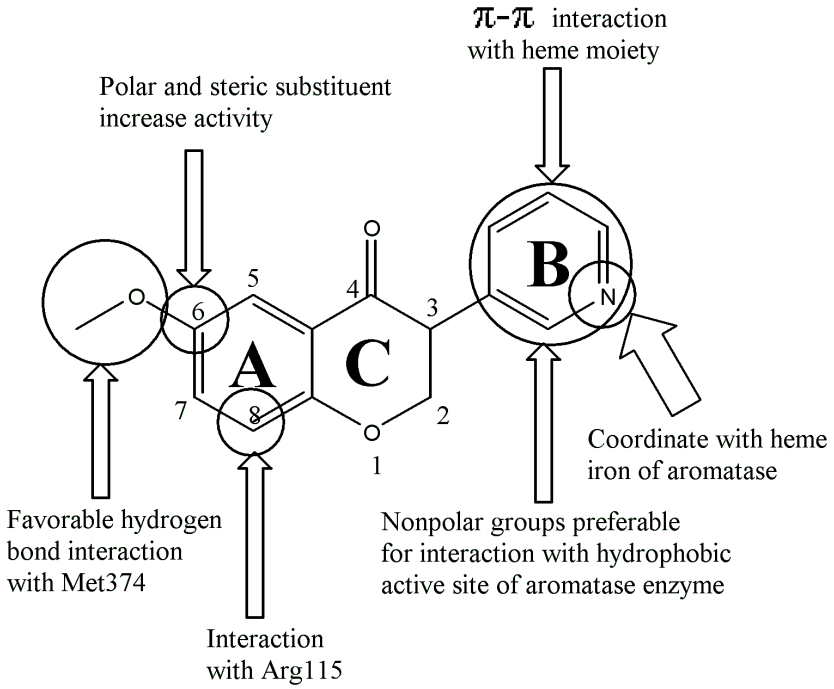
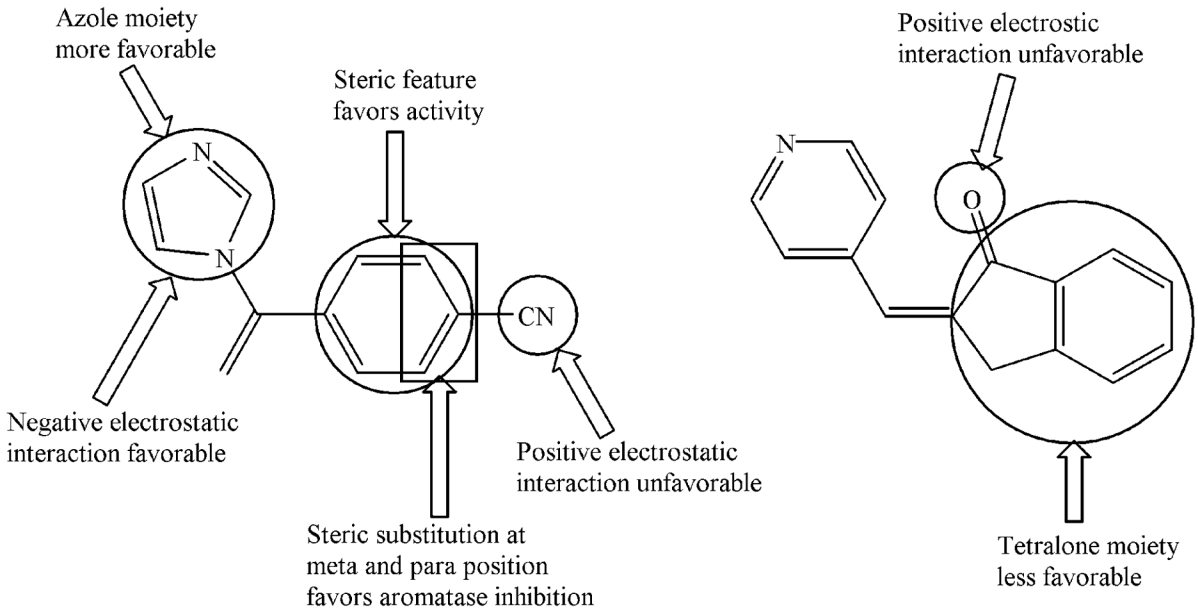


Figure 21. Structural requirements of azoles and tetralones for promising active NSAIs



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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.

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

EBSCOhost - printed on 2/14/2023 7:16 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

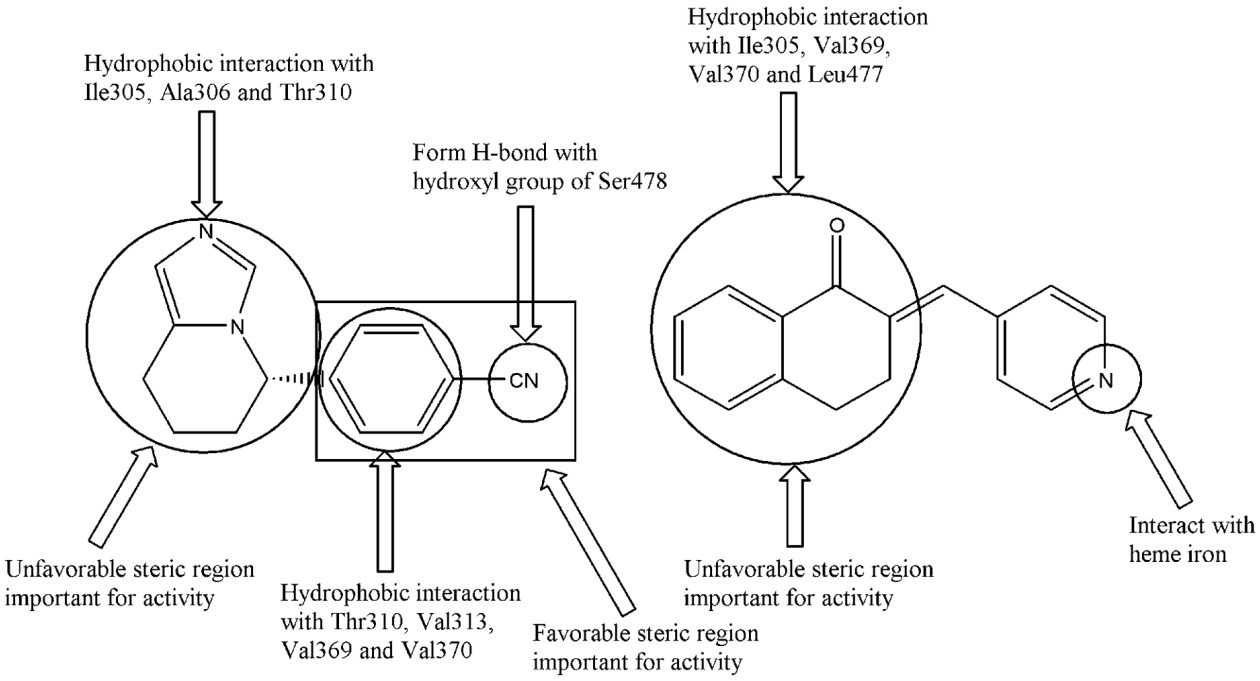
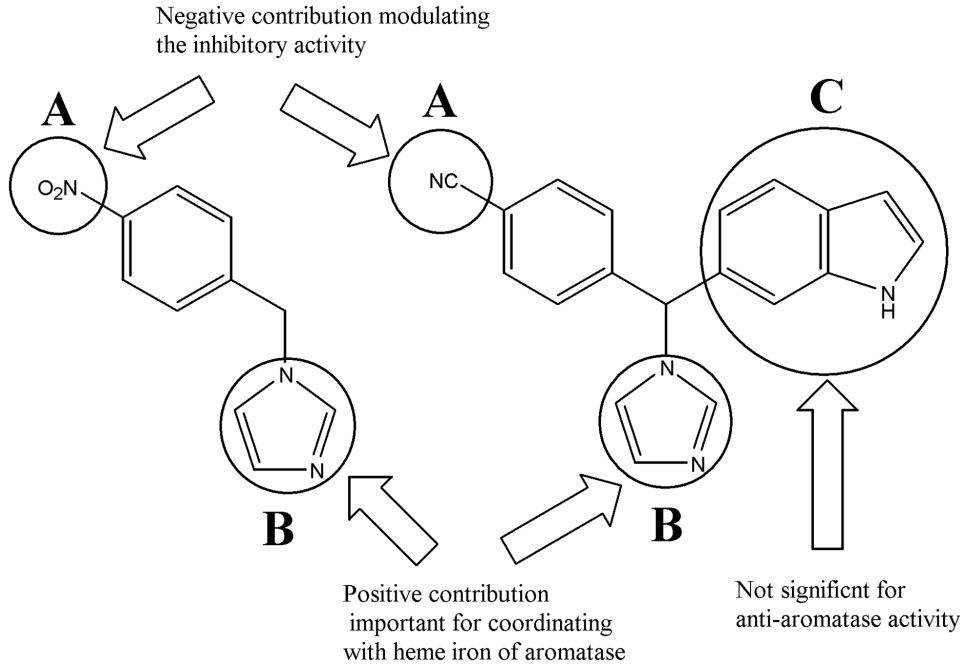


Figure 23. Important structural features required for better NSAIs



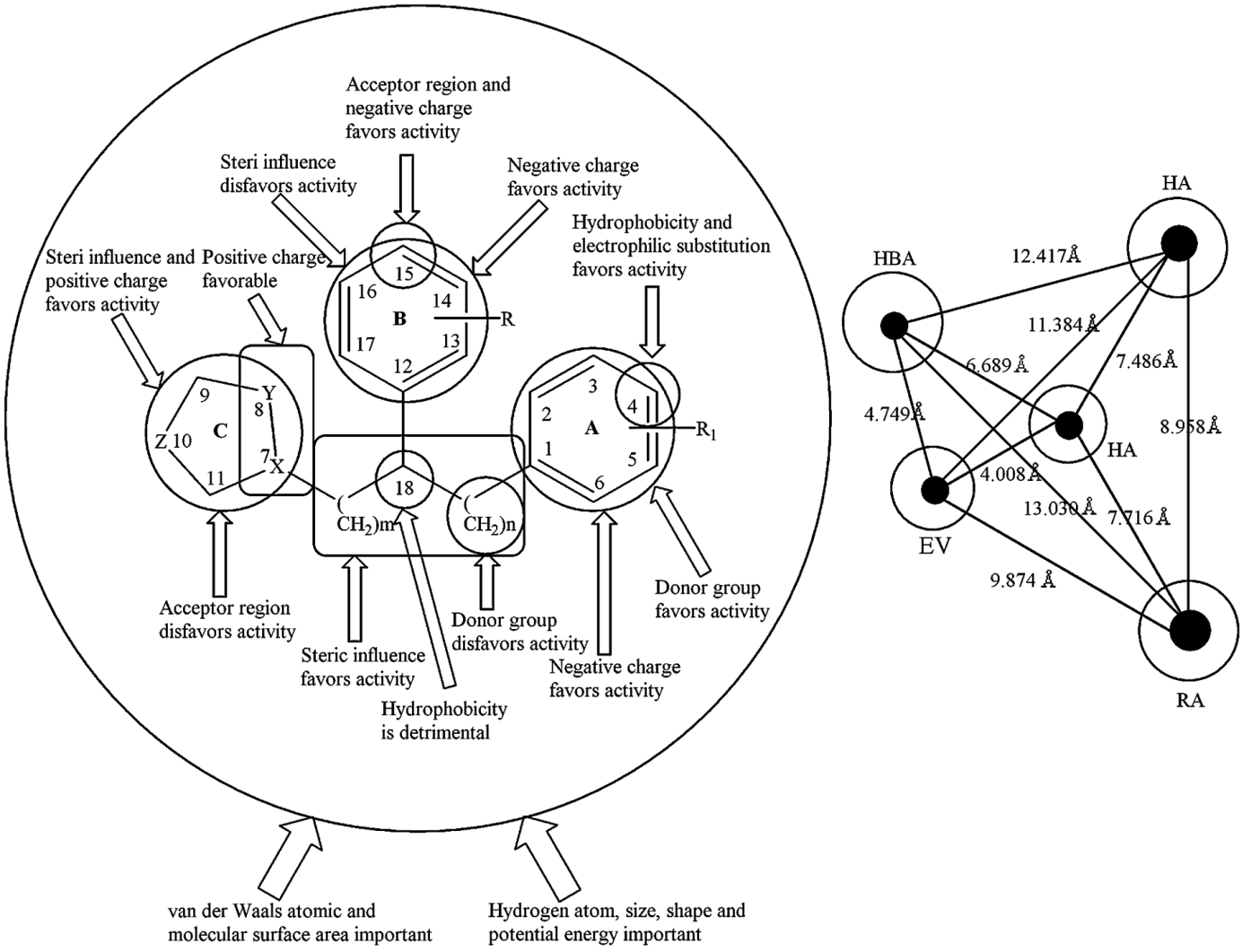
EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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 aroma-

EBSCOhost - printed on 2/14/2023 7:16 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

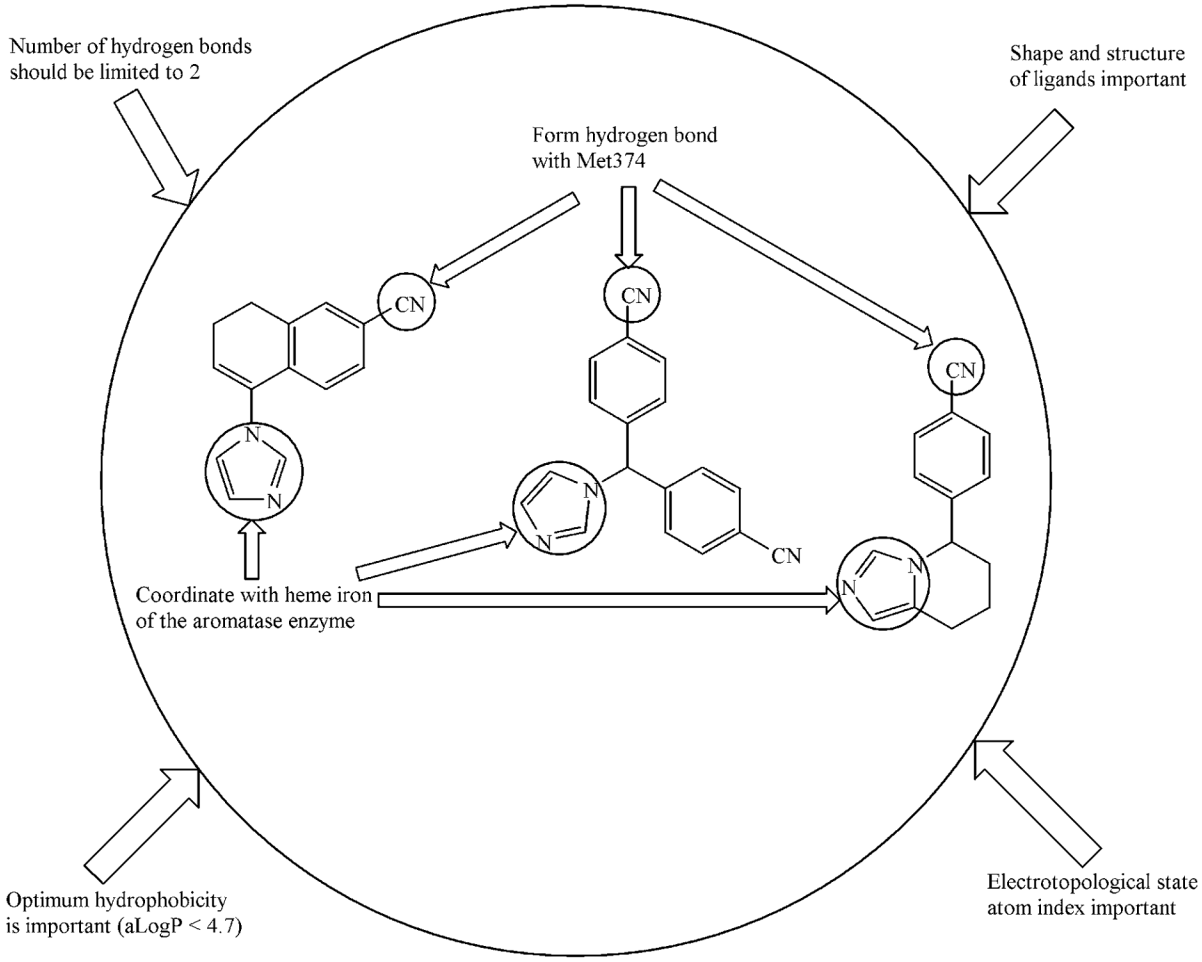


tase inhibition. Optimum hydrophobicity (aLogP < 4.7) of the ligand might play vital role for aromatase inhibition. The validated QSAR study also suggested that the cyano group containing compounds might form hydrogen bond with Met374 which is supported by the previously published result (Castellano *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

EBSCOhost - printed on 2/14/2023 7:16 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

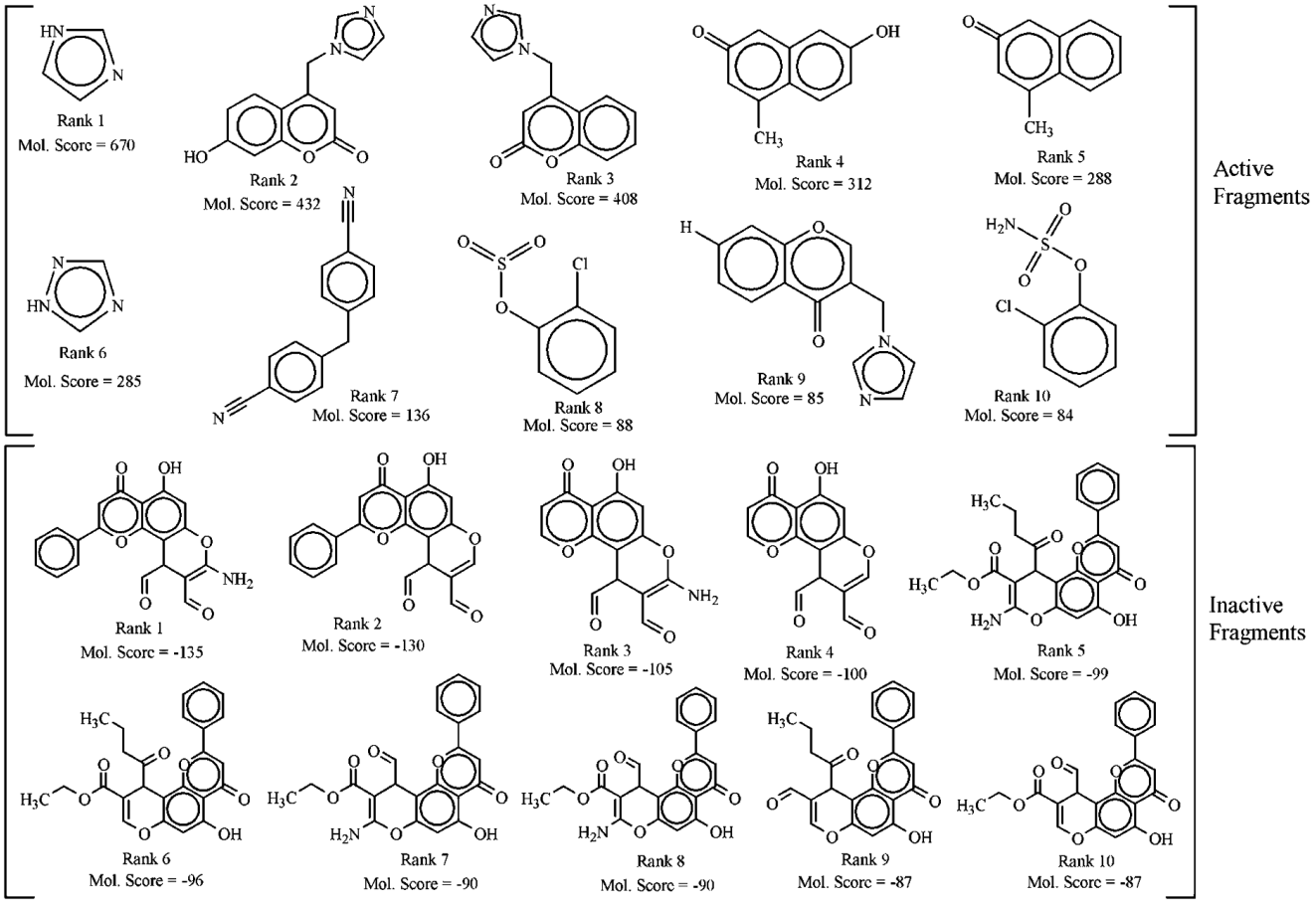


them the top ranked fragment contains 1, 3 imidazole ring system as per molecular score. Apart from that, 1, 2, 4-triazole ring system was found to be an active fragment. On the other hand, 4-[(4-cyano-phenyl) methyl] benzonitrile, a substructure of third-generation NSAI, letrozole (**6**), was found 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 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

EBSCOhost - printed on 2/14/2023 7:16 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



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.

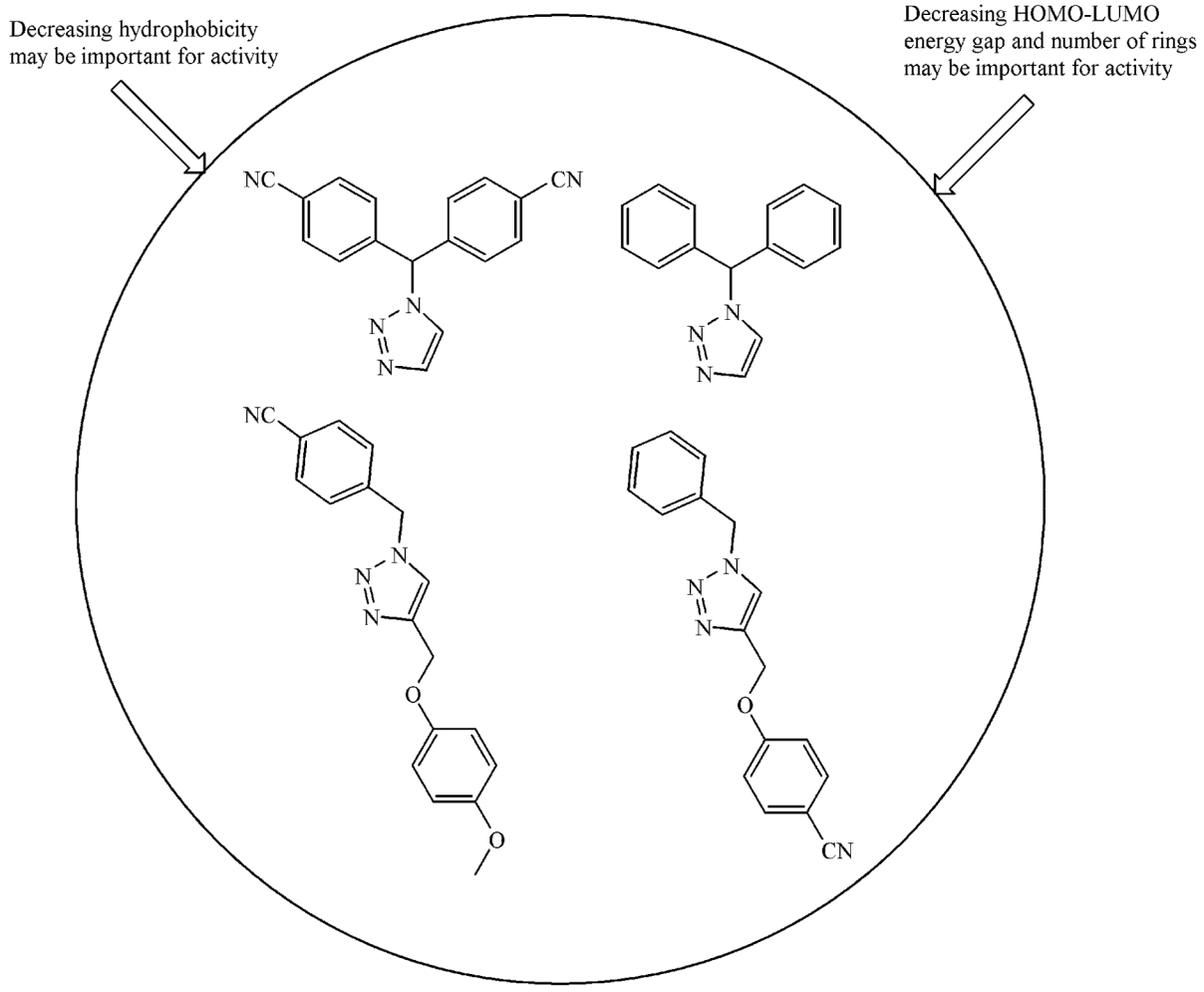
A QSAR study on 1-substituted 1, 2, 3- triazole analogs of letrozole was done (Nantasenamat *et al*., 2013b) using multiple linear regression (MLR), artificial neural network (ANN) and support vector machine (SVM). Geometry optimizations of these compounds were done by AM1 method followed by

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 hy-

Figure 27. Important structural features of 1-substituted 1, 2, 3- triazole analogs for anti-aromatase activity

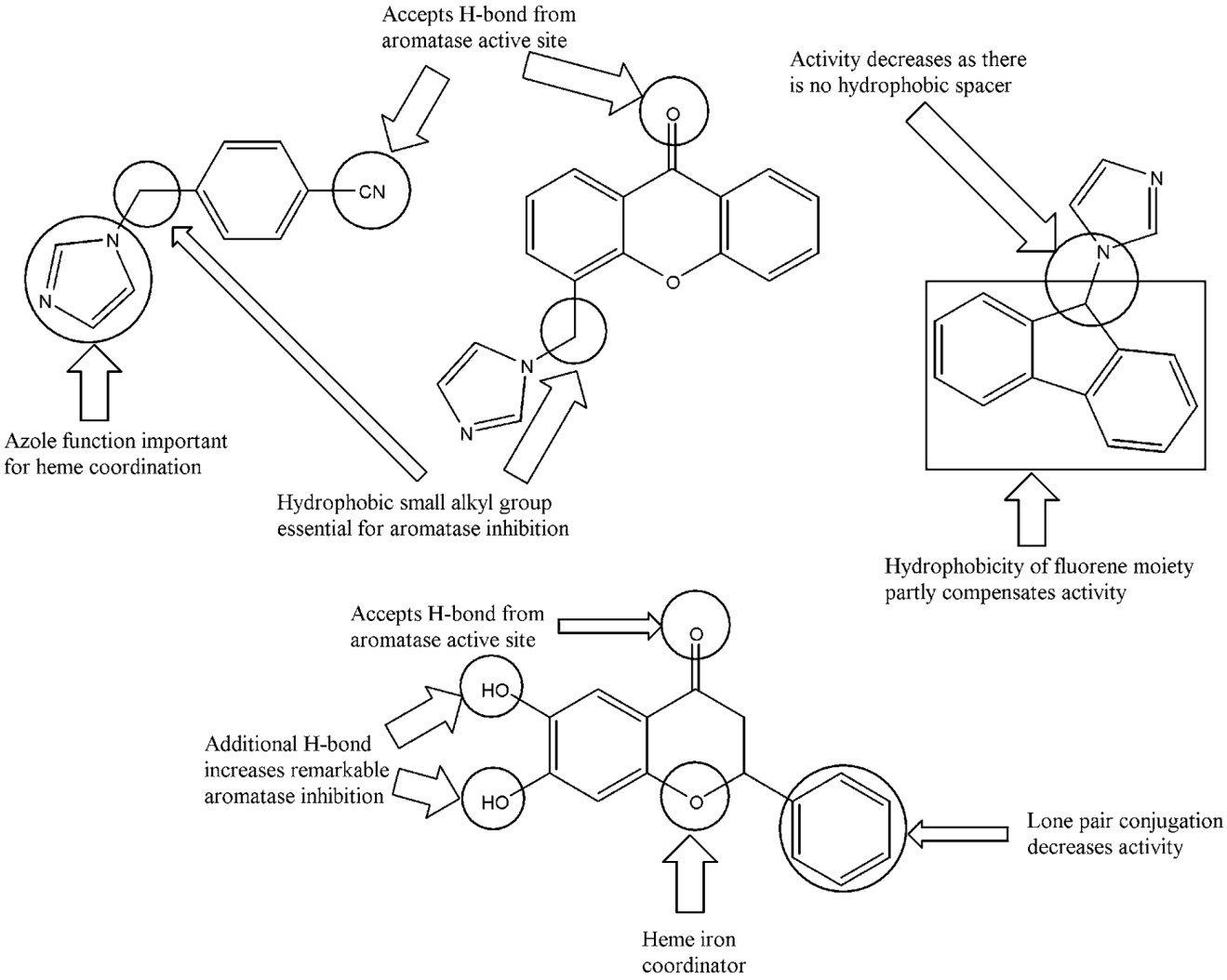


EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

droxylation. Chemical groups like CN, NO_2, CO, etc. of the NSAIs were able to accept a hydrogen bond 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

Figure 28. Structural requirements of NSAIs for better activity



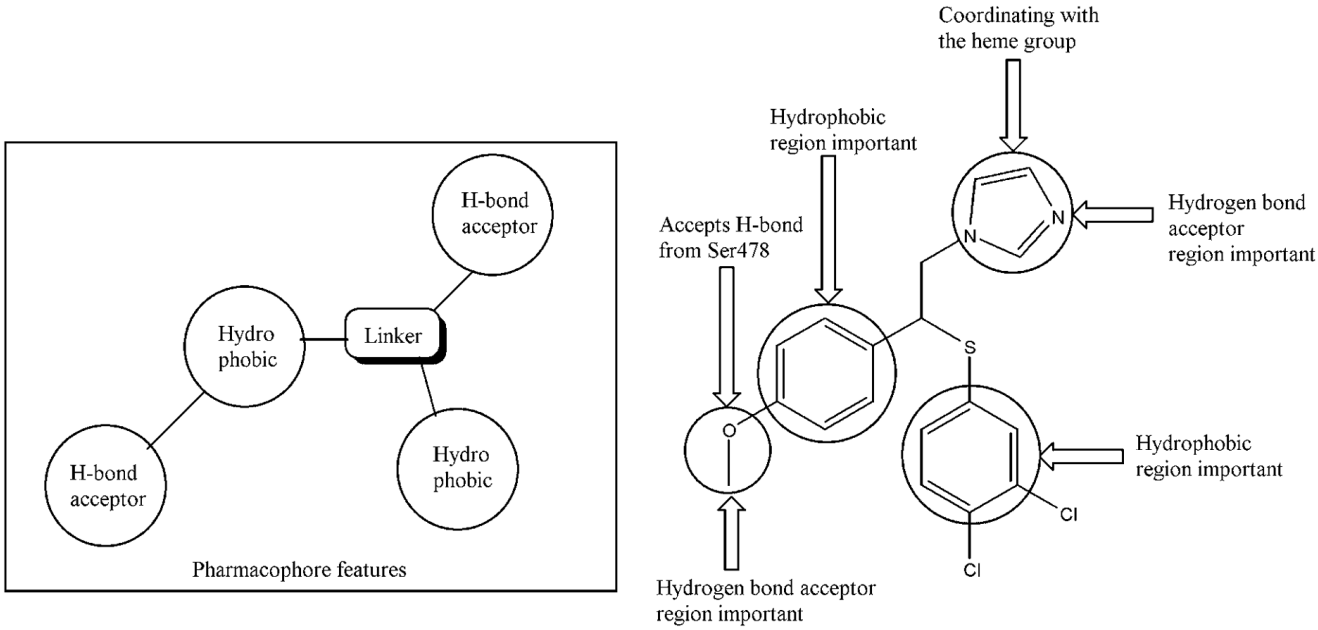
EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 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. 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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 29. Schematic representation of the pharmacophore requirements for potent NSAIs

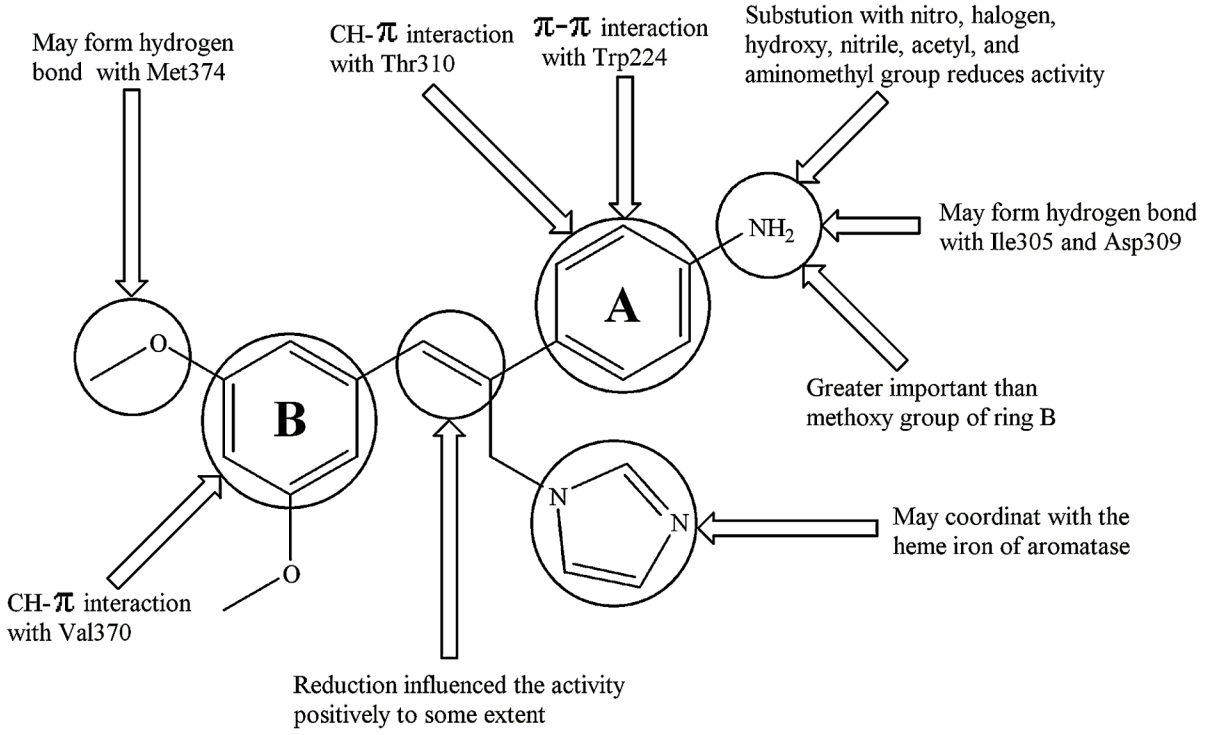


Thr310. The methoxy group of the ring B might form hydrogen bond with the backbone amino group of Met374 as well as the ring B was involved with methyl group of Val370 through CH-π interaction. This docking result is in agreement with the crystallographic data (Ghosh *et al*., 2009). These docking results 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 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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 30. Structural requirements of resveratrol analogs for anti-aromatase activity

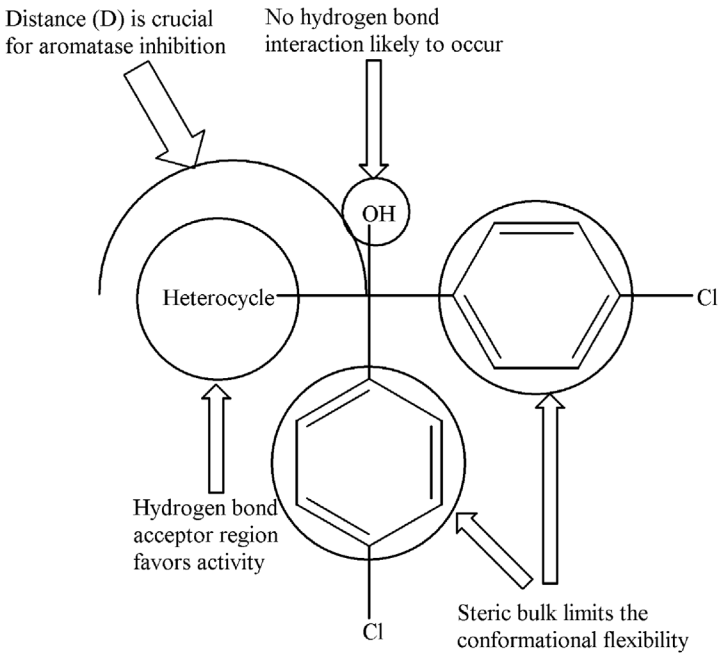


inhibitors. All these QSAR models developed showed that the distance (D) between the methane core carbon and the heterocyclic nitrogen was the crucial factor for aromatase inhibition. The steric bulk of the substituents about the chiral carbon center markedly limited conformational flexibility so that the 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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 31. Structural requirements of the dichlorophenyl compounds for NSAIs

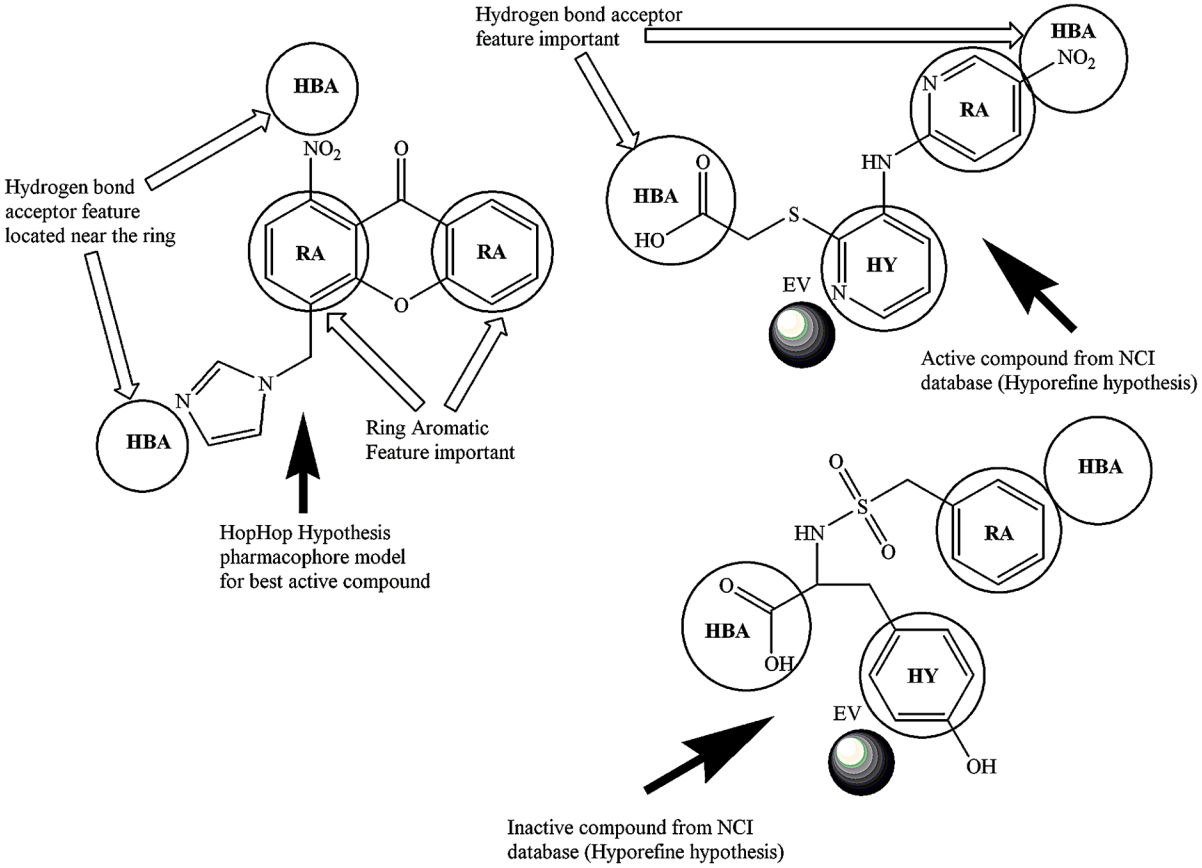


increased and two compounds were discovered. The models were used as database search queries to identify active compounds from the NCI database. The pharmacophoric requirements for better active compounds are shown in Figure 32.

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 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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 32. Important pharmacophoric features required for better NSAIs



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 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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 33. The ligand-based and structure-based pharmacophore requirements for potent NSAIs

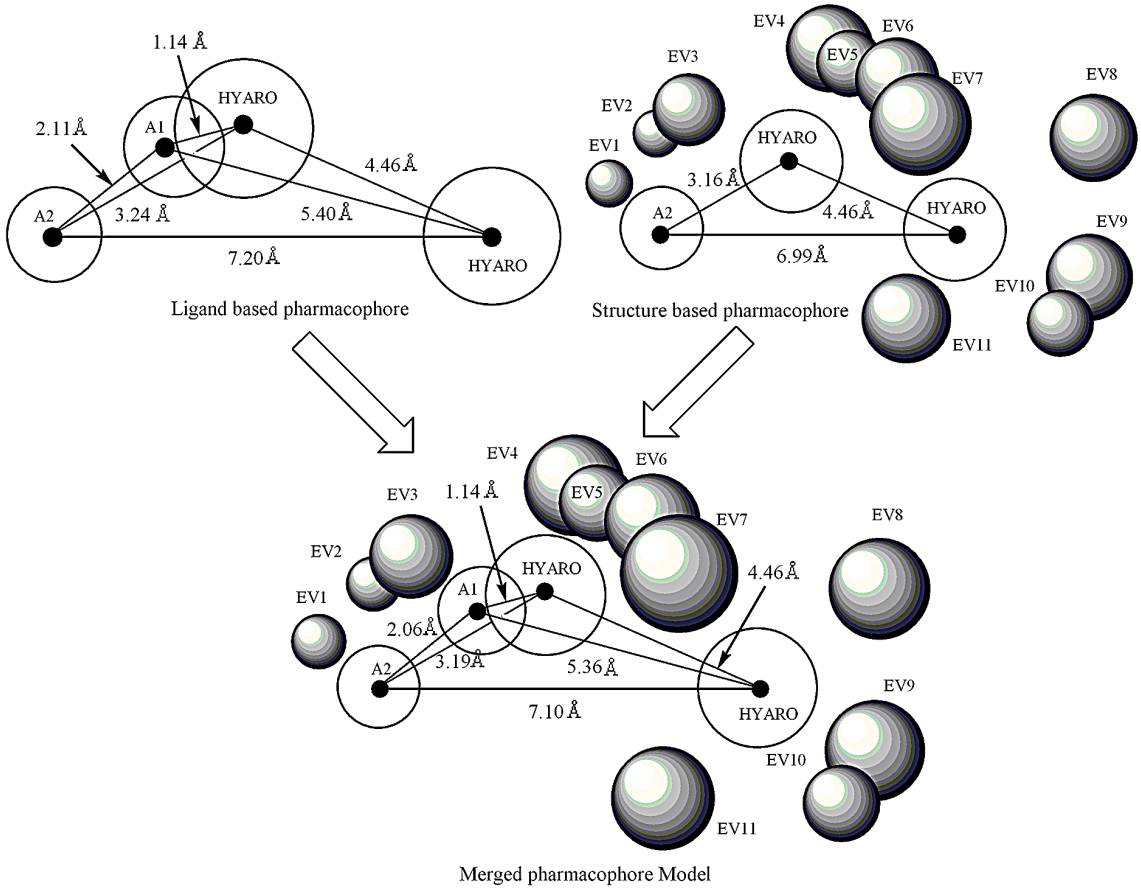
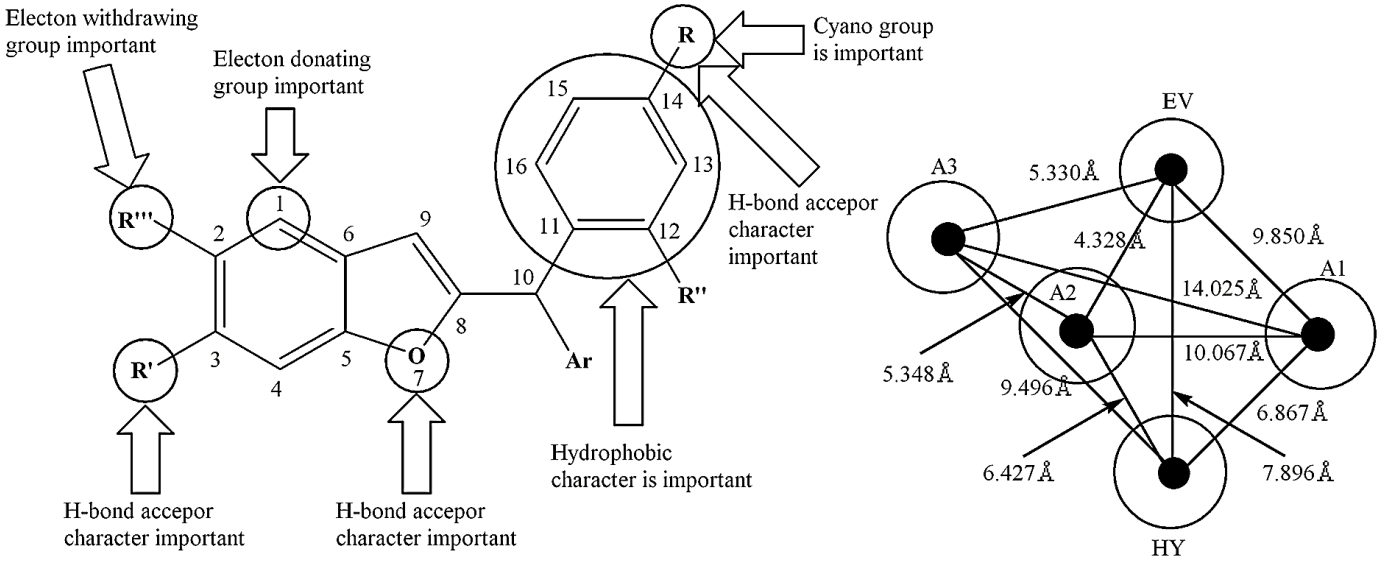


Figure 34. Structural requirements of benzofurans for potent aromatase inhibitiory activity

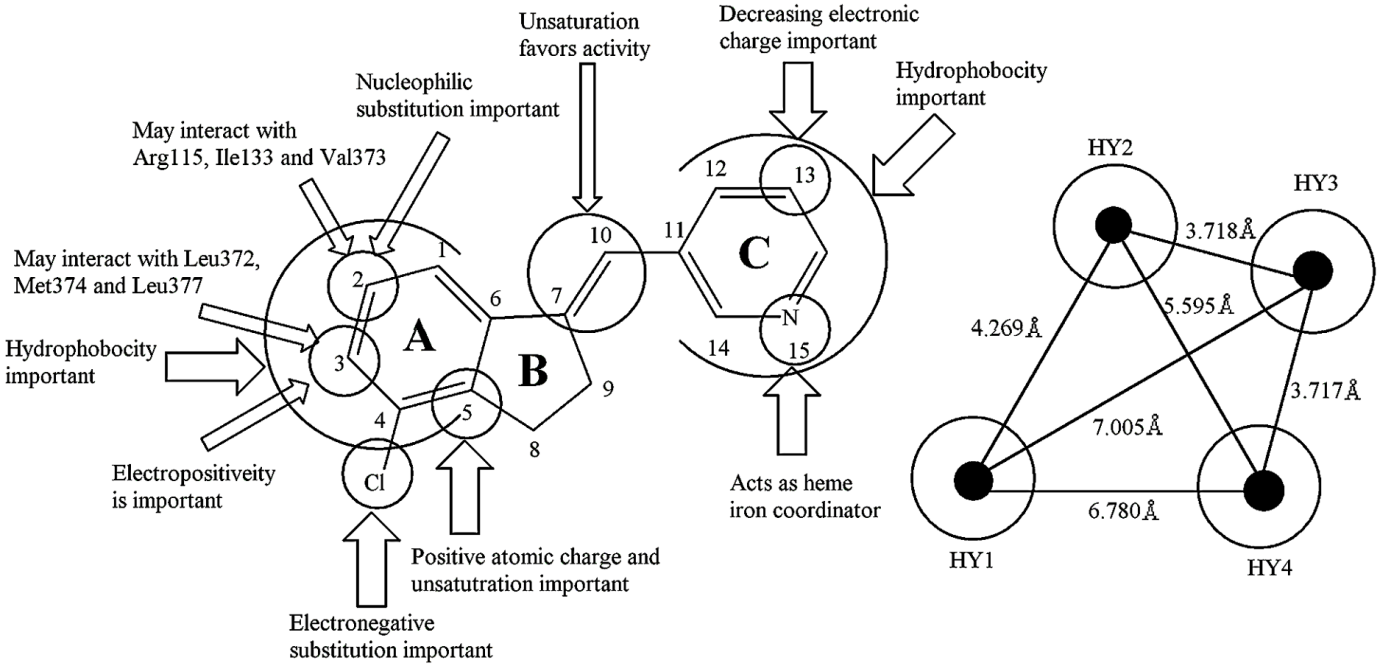


EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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 35. Structural requirements of benzocyclo derivatives for potent anti-aromatase activity



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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.

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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 36. Important structural features of natural compounds from mushroom for anti-aromatase activity

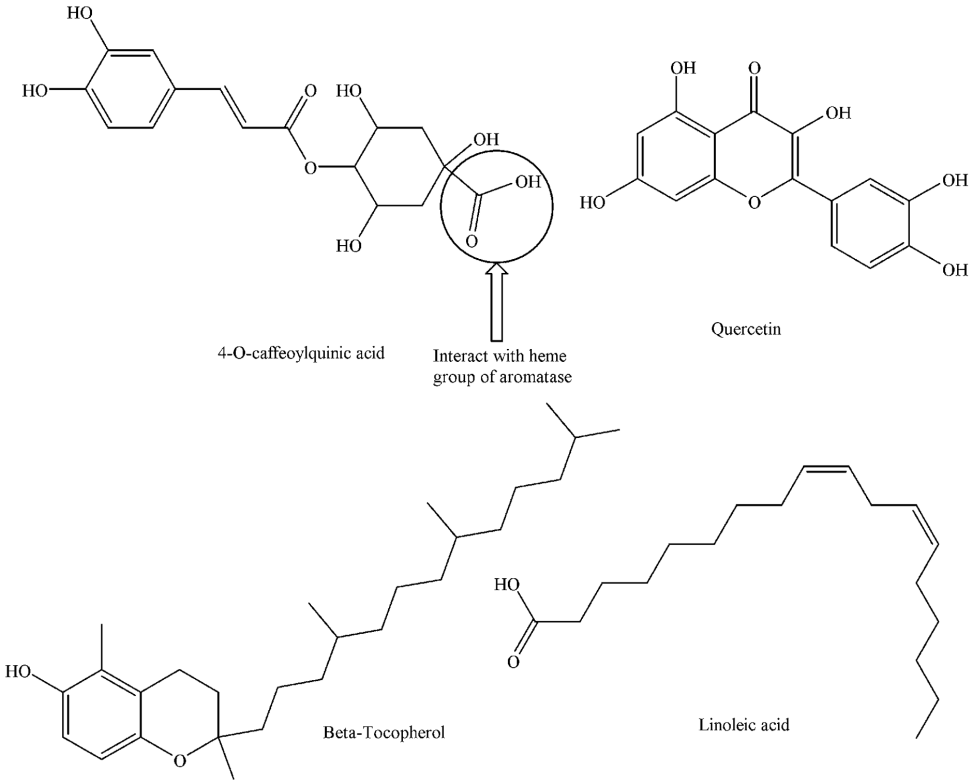
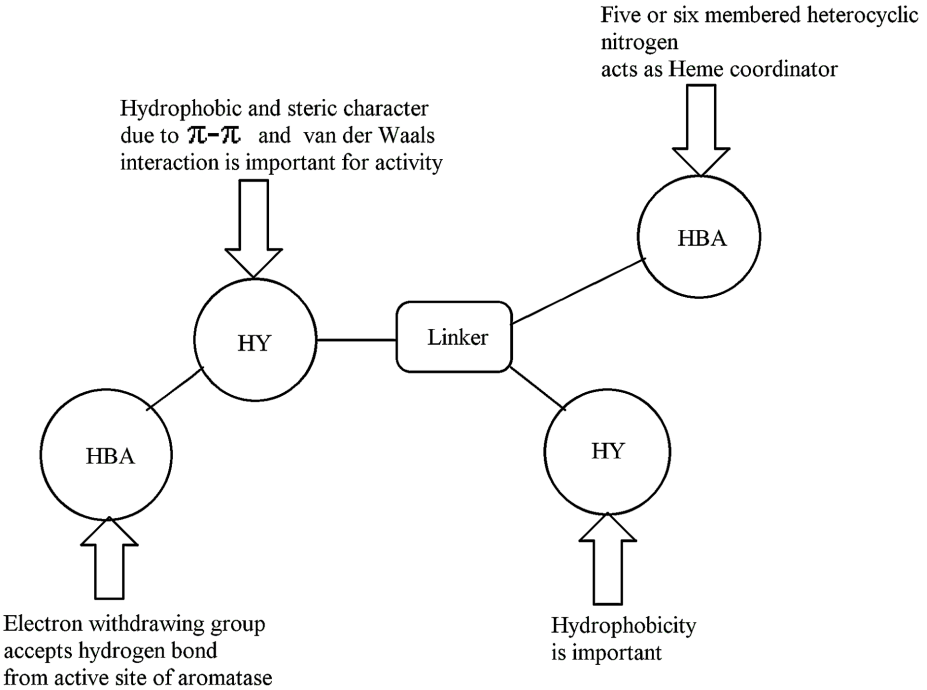


Figure 37. Schematic representation for designing potential and selective non-steroidal aromatase inhibitors



EBSCOhost - printed on 2/14/2023 7:16 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.

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 12
Computational Techniques Application in Environmental Exposure Assessment

Karolina Jagiello
University of Gdansk, Poland

Tomasz Puzyn
University of Gdansk, Poland

ABSTRACT

In this chapter, the application of computational techniques in environmental exposure assessment was described. The most important groups of these techniques are Multimedia Mass-balance (MM) modelling and Quantitative Structure-Activity/Structure-Property Relationships (QSAR/QSPR) modelling. Multimedia Mass-balance models have been widely utilized for studying Long-Range Transport Potential (LRTP) and overall persistence (P_{OV}) of Persistent Organic Pollutants (POPs), regulated by many national and international acts, including the Stockholm Convention on POPs. Recently, a novel modelling methodology that links QSPR and MM has been implemented. According to this approach, the physical/chemical properties required as the input variables for multimedia modelling can be calculated directly from appropriate QSPR models. QSPR models must be previously developed based on the relationships between the chemical structure and the modelled properties (QSPR).

INTRODUCTION

The commercialisation of new products that contain growing volumes of various chemical compounds, is the result of progress in civilisation. Certain chemicals, however, may have a negative impact on the human body and natural environment. Furthermore, noxious chemical compounds may be generated as the side products of the manufacturing processes. Consequently, development of reliable methods of risk assessment, which would allow eliminating potentially dangerous chemical substances at the stage of synthesis or final product planning, seems to be increasingly essential. Such methods should ensure good pace and cost-efficiency so as to avoid additional expenses.

DOI: 10.4018/978-1-4666-8136-1.ch012

Copyright © 2015, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

The provisions contained in the REACH Regulation (EC, 2006) that requires the entities to file every new chemical substance to be commercialised on the European market in large quantities (either manufactured or imported) with the European Chemicals Agency in Helsinki, Finland, have applied to the European Union countries since recently. The registration process must be preceded by the phase of evaluation of the chemical risk. Article 13 of the REACH Regulation determines that information on the substance should be generated, if only possible, using so called "alternative methods" with regard to ethically doubtful and expensive tests on animals. Computer assisted methods, specifically those using the quantitative modelling of relationships between the chemical structure and activity (Quantitative Structure-Activity Relationships, QSAR) and physical properties (Quantitative Structure-Property Relationships, QSPR) may be listed from among "alternative methods".

Special attention is given to the chemical substances which are defined by the acronym PBT (Persistent, Bioaccumulative and Toxic) and classified as: persistent (P), bioaccumulative (B), and toxic (T) either in the European REACH Regulation or in the legislation concerning chemical safety and applicable to Canada (CEPA, 1999), United States of America (USEPA, 1999), or Japan (METI, 1973). It should be mentioned here that the criteria of eligibility of a substance for PBT, Table 1, originate from previous research into the group of Persistent Organic Pollutants (POPs) (Gobas, 2009; van Wijk, 2009). The awareness of the risk connected with the presence of POPs in the natural environment at the turn of the 1960's and 1970's allowed launching many initiatives of multinational nature, with the aim to decrease the production and usage of such substances in order to reduce the emissions of chemical pollutants. In this context, POPs may be considered as "model pollutants" the tests of which have allowed developing many theories of transport and deposits in natural environment as well as toxicity and ecotoxicity of pollutants, such theories being largely applied these days. Recently, more and more attention is given to brominated and bromochlorinated POPs (Br-POPs and Br/Cl-POPs) given their increasing emissions to the environment and high levels of Br-POPs and Br/Cl-POPs observed in environmental matrices (Batterman, 2007; Du, 2010; Hutson, 2009), as well as their confirmed toxicity (Birnbaum, 2003).

The present chapter describes the existing scientific knowledge in two areas of concern, that is:

1. Fundamental elements of QSAR/QSPR methodologies, and
2. Development of the methodology of predicting transport of chemical substances in the environment on the basis of combination of Multimedia Mass Balance Models (MM) with QSPR models (QSPR-MM hybrid models).

BASIC IDEA AND INCREASING IMPORTANCE OF QSPR/QSAR METHODS

QSAR/QSPR models rely on a shared assumption that the variance of modelled quantity **y** (either physical/chemical property or biological activity, respectively) in a group of chemical compounds featuring similar structures is determined by the variance connected with differences in chemical structure of such compounds, whereas the mutability of chemical structure is expressed numerically by means of structural descriptors **X**. Hence, if we have experimentally measured values of a physical/chemical quantity or of activity **y** for a suitably number of compounds as well as the values of structural descriptors **X** for all the compounds, we will be able to build a suitable model in the format shown in Equation (1).

$$y = f(X). \tag{1}$$

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 1. Criteria for the identification of PBT/vPvB under annex XIII of the REACH regulation (EC, 2006)

	Persistence (P)	Bioaccumulation (B)	Toxicity (T)
PBT Substances	t1/2 in marine water > 60 days or t1/2 in fresh / estuarine water > 40 days or t1/2 in marine sediment > 180 days or t1/2 in fresh / estuarine sediment > 120 days or t1/2 in soil > 120 days	BCF > 2000	NOEC (long-term) < 0.01 mg/L • Substance is classified as carcinogenic (category 1 or 2), mutagenic (category 1 or 2), or toxic for reproduction (category 1, 2 or 3), or • There is other evidence of chronic toxicity, as identified by the classifications: T, R48, or Xn, R48 according to Directive 67/548/EEC
vPvB Substances	t1/2 in marine, fresh/estuarine water > 60 days or t1/2 in marine, fresh/estuarine sediment > 180 days or t1/2 in soil > 180 days	BCF > 5000	N/A

The formula allows predicting the modelled quantity for the remaining compounds from the group, on the basis of their structural descriptors (Cronin, 2010). In QSAR/QSPR, the modelling procedure is composed of three stages:

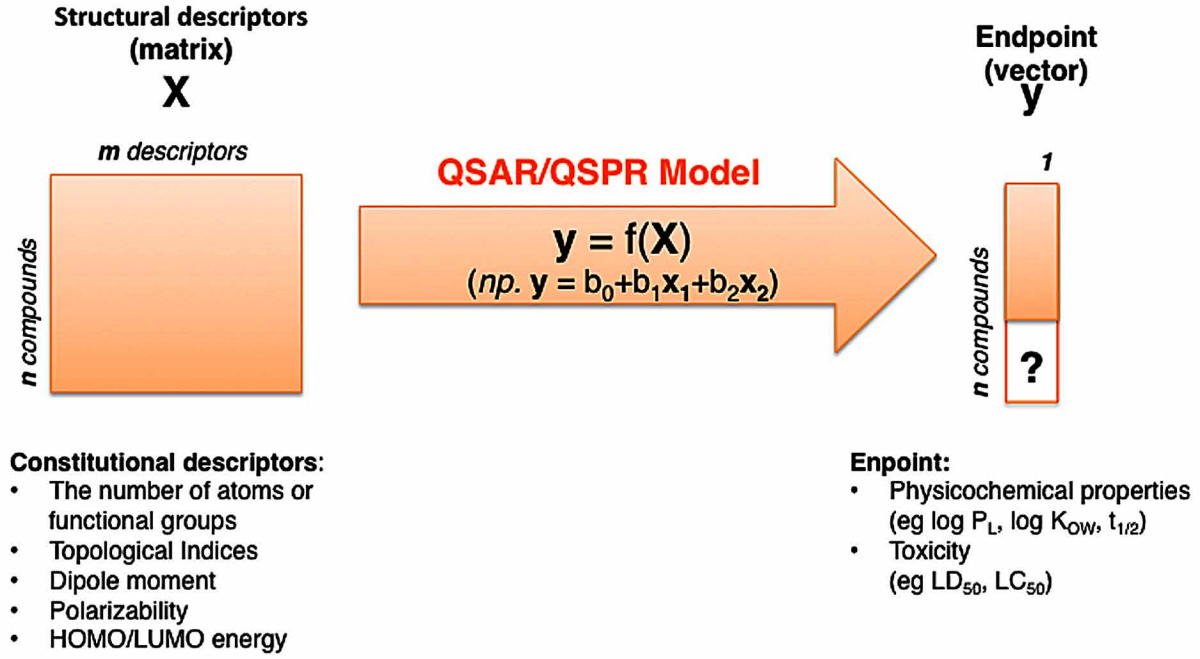
1. Model calibration on the basis of a part of compounds assigned to a "training set";
2. Model validation which accounts for test of model functioning in the case of compounds which are not covered by the process of calibration and which have a defined quantity **y** (compounds from the validation set); and
3. Model application for predicting the quantity **y** in the case of new compounds (Figure 1) (Cronin, 2010; Puzyn, 2009a).

In compliance with the recommendations of validation (OECD, 2007) adopted in OECD countries, QSAR/QSPR model should feature the following characteristics:

1. Well defined modelled quantity **y** (endpoint);
2. An unambiguous algorithm which allows repeating the tests;
3. Defined applicability domain of the model, which means a very well-defined group of chemical compounds for which quantity **y** is predicted by the model as a result of interpolation and hence, the value of the endpoint may be deemed correct;
4. Appropriately used and interpreted statistical quantities which describe the goodness-of-fit, flexibility (robustness) and predictive ability of the model; and
5. Physical interpretation of applied combination of structural descriptors, if possible.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 1. Scheme of QSAR/QSPR models development



If the model meets all the five criteria, it may be used as a reliable tool in the procedure of new chemical compound risk assessment.

QSPR/QSAR MODELLING ALGORITHM

QSPR/QSAR modelling procedure is composed of three stages:

1. Model calibration;
2. Model validation, and
3. Model application for the purpose of predicting the quantity **y** for new compounds (Cronin, 2010; Puzyn, 2009b).

Model Calibration

The first stage is initiated by the construction of molecular models of all the molecules of a group, and is followed by the computation of the matrix of structural descriptors **X** on the basis of such molecular models. The matrix contains (in columns) the descriptors for each of the compounds (matrix rows) (Cronin, 2010; Puzyn, 2009b).

The values of modelled **y** are simultaneously collected (or measures) for as large part of compounds from the group as possible. Similarly as in the case of each mathematical model, QSPR/QSAR modelling will never lead to predictions on the quality higher (smaller error) than the quality of data used for such

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

model calibration. Occurrence of additional variability sources that are the result of different conditions (e.g. temperature and pressure) of the experimental study, laboratory-specific differences in the individual studies, another methodology applied for the purpose of determination of the same quantity, etc., should be as minimal as possible. If not, the modelling process will apply firstly to the impact of such other features rather than to the relationship between the quantity **y** and the chemical structure (Puzyn, 2009b; Cronin & Schultz, 2003; Dearden, 2009).

If matrix **X** and vector **y** are available, the construction of QSPR/QSAR model will commence with three chemometric methods being the most commonly used. These are: MLR method (Multiple Linear Regression) as well as two methods employing the latent vectors which are linear combinations of original descriptors, that is: PCR method (Principal Component Regression) and PLS method (Partial Least Squares) (Puzyn, 2009b; Walczak, 2009). The complexity and the type of actual method should depend on the specificity of the problem, which means that the simplest method should be always applied, whenever possible. The two last methods will be explicitly useful if individual descriptors in matrix **X** demonstrate high inter-correlation (Puzyn, 2009a).

Furthermore, selection of the optimum combination of descriptors for the purpose of modelling seems to be an important issue. Two alternative approaches to the problems are used, conditional on the availability of information connected with the mechanism which decides on the specificity of either structure-property or structure-activity relationship. If the mechanism and structural factors having an impact on quantity **y** are known, the descriptors might be selected arbitrarily. If not, special algorithms need to be employed in order to test thousands of descriptor combinations and select such combinations that lead to the model featuring the lowest-error-possible predictivity (Puzyn, 2009b; Gramatica, 2010). Final selection, however, must be made by the researcher responsible for construction of the model, who takes into account the possibility of assignment of the physical sense to the combination of descriptors as proposed by the application. Moreover, selection of descriptors for a model should take into account the fact that excessive number of descriptors (variables) compared with the number of compounds used for the purpose of calibration almost certainly leads to excessive model matching (overfitting) and hence, to the loss of information generalisation capacity and of correct predictions of variable **y** for new compounds. Empirical studies allow making an assumption that at least 5 calibration compounds should be assigned to a single descriptor (Topliss ratio) as a standard in QSPR/QSAR modelling (Gramatica, 2007, 2010; Holland, 1992; Livingstone & Salt 2005; Topliss & Edwards, 1979).

Model Validation

The second stage of QSPR/QSAR modelling consists in a reliable validation. The validation contains the evaluation of goodness-of-fit, evaluation of model robustness, evaluation of predictive capacity in case of new compounds (predictivity), and verification of the model applicability domain (Puzyn, 2009b; Gramatica, 2010).

Evaluation of model goodness-of-fit is conducted on the basis of values of residuals for the compounds used for calibration. Distribution of residuals should be similar to the normal (random) distribution. The most frequently used measurements of ability of QSAR models to reproduce the training data are determination coefficient R^2 (2) and $RMSE_C$ (root mean square error of calibration) (3):

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i^{obs} - y_i^{pred})^2}{\sum_{i=1}^{n}(y_i^{obs} - y^{obs})^2} \tag{2}$$

$$RMSEC = \sqrt{\frac{\sum_{i=1}^{n}(y_i^{obs} - y_i^{pred})^2}{n}} \tag{3}$$

where: y_i^{obs} – experimentally determined value of modelled quantity for the *i*-th calibration compound; y_i^{pred} – model-predicted value of modelled quantity for the *i*-th calibration compound, and n – number of compounds used for model calibration (Puzyn, 2009b; Gramatica, 2007).

Model robustness describes its resistance to minor changes in the set of compounds used for calibration. It is tested by means of one-by-one exclusion of individual elements (or groups of 'm' elements) from the set and re-calibration for n-1 (or n-m, respectively) compounds. Such model is the basis for predicting the value of modelled quantity for an excluded element (or excluded elements). Next, the excluded element (or excluded elements) is returned to the training set and the entire process is repeated on the next parts of the training set. The process is known as "cross-validation" method. If a single element is excluded in each cycle, the process is named a "leave-one-out" type cross-validation; if a larger part of the set is excluded once in a cycle, the process is called a " leave-many-out" type cross-validation. Validation statistics, i.e. cross-validation coefficient Q^2_{CV} (4), and RMECV (root mean square error of cross-validation) (5) are calculated on the basis of the values of residuals which, in turn, are calculated as differences between experimentally determined modelled quantity values ($\mathbf{y^{obs}}$) and model-predicted values for one-by-one excluded elements ($\mathbf{y^{pred_cv}}$):

$$Q^2_{CV} = 1 - \frac{\sum_{i=1}^{n}(y_i^{obs} - y_i^{pred_cv})^2}{\sum_{i=1}^{n}(y_i^{obs} - \bar{y}^{obs})^2} \tag{4}$$

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{n}(y_i^{obs} - y_i^{pred_cv})^2}{n}} \tag{5}$$

where: $y_i^{pred_cv}$ means the model-predicted value of modelled quantity for the i-th compound excluded from the training set during the cross-validation. The term "cross-validation" is also known as internal model validation since the same elements (although not simultaneously) are used for model calibration and validation (Puzyn, 2009b; Gramatica, 2007).

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

On the other hand, the external validation is the only reliable test that confirms that the resultant model will correctly predict the values **y** for 'k' new chemical compounds. External validation relies on prediction of modelled quantity ($\mathbf{y}^{\mathbf{pred}}$) for such chemical compounds that have not been used in the process of calibration. Consequently, the correctness of prediction may be easily verified on the basis of observed differences (residuals) between the values $\mathbf{y}^{\mathbf{pred}}$ and $\mathbf{y}^{\mathbf{obs}}$. In this case too, the external validation coefficient (6) and RMSEP (root mean square error of prediction) will be calculated (7):

$$Q^2_{Ext} = 1 - \frac{\sum_{j=1}^{k}(y_j^{obs} - y_j^{pred})^2}{\sum_{j=1}^{k}(y_j^{obs} - \bar{y}^{obs})^2} \tag{6}$$

$$RMSEP = \sqrt{\frac{\sum_{j=1}^{k}(y_j^{obs} - y_j^{pred})^2}{k}} \tag{7}$$

where: y_j^{obs} – experimentally determined value of modelled quantity for *j*-th validation compound; y_j^{pred} – model-predicted value of modelled quantity for *j*-th validation compound; k – number of compounds used for the purpose of external validation of the model (Puzyn, 2009b; Gramatica, 2007).

R^2, Q^2_{CV}, and Q^2_{Ext} are interpreted as coefficients of correlation between observed and predicted values **y** and describe, which part of vector **y** variation is explained by the model. Coefficients R^2, Q^2_{CV} and Q^2_{Ext} of QSPR/QSAR model should be almost equal to one, if possible, whereas the differences between these coefficients (e.g. $R^2 >> Q^2_{Ext}$) may point out excessive matching between the model and the calibration points and the resultant reduction in predictability of **y** for new compounds (overfitting). Similarly, the values of RMSEC, RMSECV, and RMSEP errors in a defined model should be close to each other and obviously, the smallest possible (Puzyn, 2009b; Gramatica, 2007). Measurements described above are the most frequently used ones. Description of other very useful validation parameters, such as concordance correlation coefficient (CCC), and "r^2_m" metrics can be found in Chirico and Gramatica (2012) and Roy (2013), respectively.

The validation stage includes also characterisation of the applicability domain. This term (applicability domain) describes a theoretical space determined by the values **X** and **y**, within which the predictions are reliable. Hence, one must check for which types of compounds, due to their similarity to the training set, the predictions will be reliable and for which compounds a defined model will prove irrelevant. Detailed review of the methods designed for definition of the boundaries of the applicability domain may be found in papers by Netzeva (2005) and Eriksson (2003).

Only validated QSAR/QSPR model can be apply to predict modelled valuables for compounds for which these data are not available.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

EVALUATION OF MOBILITY OF POPS IN THE ENVIRONMENT

Effective Predicting of Congener Characteristics

From among twenty four POPs covered by the Protocol on Persistent Organic Pollutants to the 1979 Convention on Long-Range Transboundary Air Pollution (Aarhus Protocol) (UN-ECE, 1998) and from among twenty one POPs covered by the provisions of the Stockholm Convention on Persistent Organic Pollutants (UNEP, 2001), most compounds occur in the form of families of congeners. Congeners of POPs are chemical compounds the structure of which relies on the same carbon skeleton; they differ in the number of halogen substituents, substitution patterns, and, sometimes, also in the type of halogen atoms in the molecule.

The families of congeners may be exemplified by: PCBs (209 congeners); PCNs (75 congeners); PCDDs (75 congeners); PCDFs (75 congeners), and PBDEs (209 congeners) (Figure 2). This means that in chemical context the provisions of the Aarhus Protocol and Stockholm Convention apply to a considerably higher number of single compounds, which may differ in their properties and/or toxicity given their different structures. For example, only twelve PCBs congeners ("non-*ortho* PCB" and "mono-*ortho* PCBs") as well as seven PCDDs congeners proved to have the capacity to build a complex with aryl-hydrocarbon receptor ("Ah receptor") and consequently, to induce toxic response which is specific to 2,3,7,8-dibenzo-*p*-dioxin (TCDD) (Falandysz, 1999).

Given a considerable plurality of probable congeners, determination of physical-chemical properties and/or activity for all the compounds classified into a defined family would prove very expensive and time-consuming. Besides, considerable number of experimental animals would be required for toxicity surveys. Economic and often ethical motives make the use of QSPR/QSAR computer-aided methods preferable in such cases. Hence, the development of effective methods and tools that facilitate the process of modelling large sets (families) of chemical compounds featuring similar structures seems necessary.

The most time-consuming phases of modelling the congener family properties by means of QSPR/QSAR methods are:

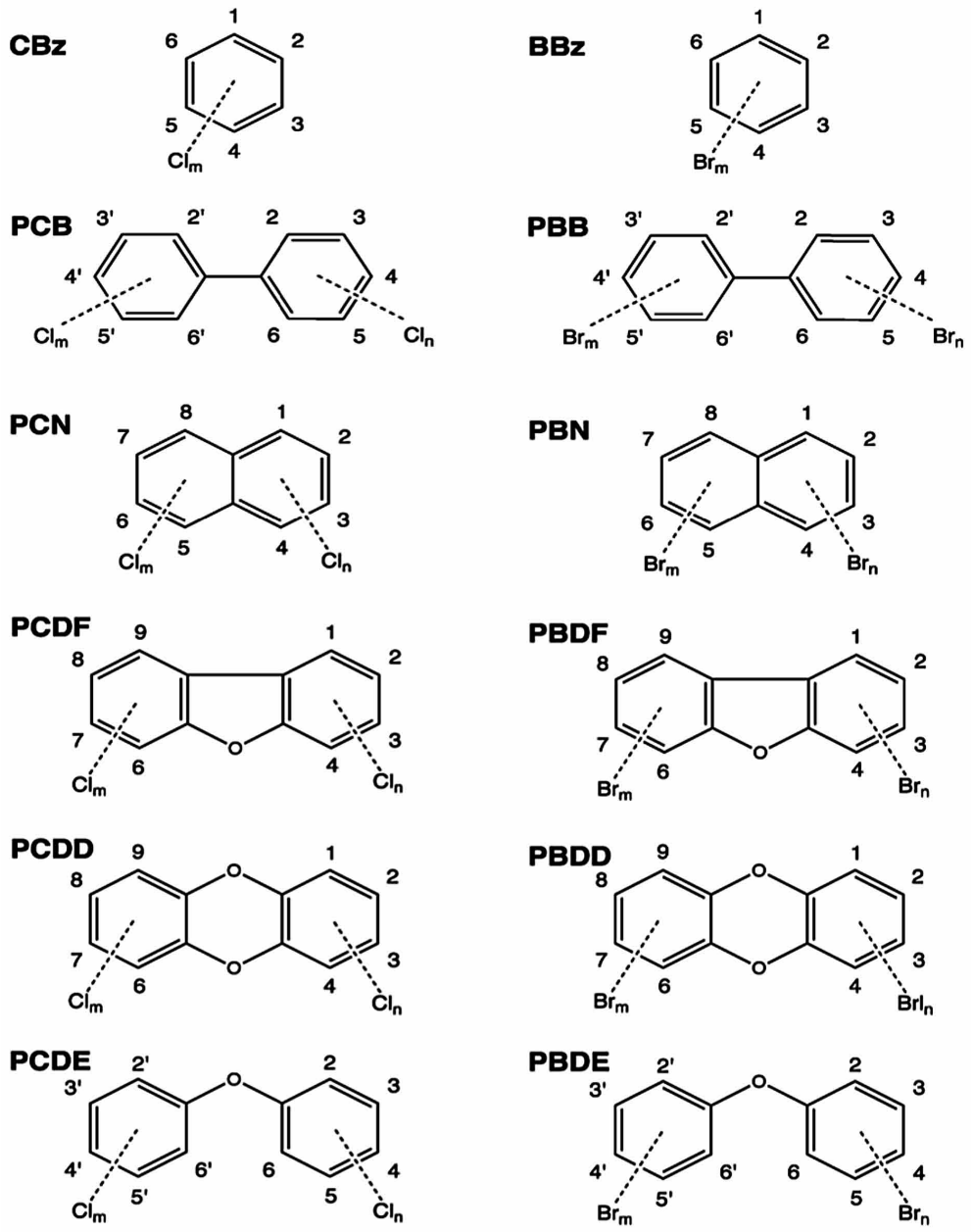
1. Generation of matrices containing the descriptions of chemical structures of individual representatives of a defined family, in the internal coordinates format (Z-matrix), and
2. Optimisation of geometry of each structure by means of suitable quantum chemistry method, which is necessary to calculate the descriptors of the structure **X**.

The "ConGENER" software may considerably reduce the timing of this phase. It will generate a series of internal co-ordinate matrices for all probable congeners on the basis of a mother (other than substituted) molecule internal coordinate matrix which indicates hydrogen atoms likely to be substituted (Haranczyk, 2008).

Haranczyk (2008) identified 287 congeners of polychlorinated pyrenes (PCPYs) with ConGENER software and then analysed their arrangement in the space of calculated descriptors by means of PCA (Principal Component Analysis). Four-factor PCA model explained 97% of the total variability within the matrix **X**. The first principal component described the major part of the variance (76%) and indicated the groups of congeners containing the same number of chlorine atoms in the molecule (mono-

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 2. Chemical structures of congener: polychlorinated benzenes (CBzs), polybrominated benzenes (BBzs), polychlorinated biphenyls (PCBs), polybrominated biphenyls (PBBs), polychlorinated naphthalenes (PCNs), polybrominated naphthalenes (PBNs), polychlorinated dibenzofurans (PCDFs), polibromowane dibenzofurans (PBDFs), polychlorinated dibenzo-p-dioxins (PCDDs), polybrominated dibenzo-p-dioxins (PBDDs), polychlorinated dipheny ethers (PCDEs), polybrominated dipheny ethers (PBDEs)



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

CPYs; di-CPYs; tri-CPYs, etc.). The second principal component (10% of the variance) indicated the compounds following the value of ionisation potential that depended on the method of chlorine atom substitution pattern. The analysis of compound arrangement in the chemical space allows concluding that the congeners that have unsubstituted hydrogen atoms in 1, 3, 6, and 8 positions should demonstrate higher resistance to oxygenation processes compared with the remaining compounds from the PCPYs group. The third principal component (7% of the variance) described the arrangement of partial Mulliken charges in the molecule. Interestingly, decachloropyrene (perchlorinated pyrene) demonstrated quite different distribution of charges compared with all the remaining PCPYs congeners. Additional calculations concerning the families of chloronaphthalenes, anthracenes, phenanthrenes, and chrysenes displayed certain regularity: arrangement of charges in the molecules of perchlorinated congeners was considerably different from the arrangement of charges in the molecules of other compounds from a defined family. Comparison with the results produced at DFT level (B3LYP functional), however, did not confirm the differences observed in the semi-empirical PM6 method. This may indicate certain obsolescence and the need to carry out additional parameterisation of PM6 methods for perchlorinated aromatic compounds. The fourth principal component (explaining 4% of the variance) distinguished the congeners by the differences in molecule polarity (their dipole moment), such differences resulting from different symmetry of substituent arrangement. The molecules with symmetrical arrangement of substituents in the individual rings (e.g. 1,3,6,8-tetrachloropyrene) feature lower values of the dipole moment compared with the congeners the molecules of which contain the substituents occurring in the neighbouring positions (e.g. 1,8,9,10-tetrachloropyrene). That method allowed identifying the main sources of structural variability in PCPYs congeners, one of the families of persistent organic pollutants (Haranczyk, 2008).

According to the increasing importance of brominated and bromochlorinated congeners, this groups of compounds were also included in the studies (Haranczyk, 2008). Introduction of additional type of substituent into the system will considerably increase the number of probable substitution combinations and the number of probable Br/Cl-POPs congeners, accordingly. Moreover, the number of congeners depends on the symmetry of the parent molecule. For example, the number of probable congeners being the derivatives of naphthalene will be considerably lower due to higher symmetry of naphthalene molecule, compared with the number of congeners being derivatives of biphenylene oxide. Application of ConGENER software package allowed calculating the number of all probable bromochlorinated congeners from individual POPs families (Table 2). Considering the fact that POPs chlorine and bromine derivatives may transform into equally and even more dangerous bromo-chlorinated derivatives as a consequence of ordinary processes occur in the natural environment, the results of the mentioned studies have shown the magnitude of the need to evaluate the risks presented by such compounds. Application of QSPR methods seems to be the only option that allows estimating the pathways of such high a number of compounds in the natural environment.

Efficient modelling of big families of compounds from POPs group requires the use of suitable method of molecular geometry optimisation as well as the calculation of structural descriptors for individual congeners. Such method should lead to reliable results and ensure the minimum time consumption possible (considerably shorter than in the case of experimental studies). To achieve this, a large part of QSAR/QSPR models published in 2004 - 2007 employed the DFT (Density Functional Theory) methods, which mostly used the B3LYP functional (Puzyn & Falandysz, 2007; Puzyn, 2007). Compared with Hartree-Fock (*ab initio*) method relying on the single-electron approximation, the DFT methods, which partly include electron correlation, produce better estimates of molecule structural parameters. Compared

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 2. Number of congeners (Haranczyk, 2008)

Compound	Number of Congeners	
	One Substituent	Two Substituents
Benzene	12	91
Naphthalene, Biphenylene, Dibenzo-p-dioxin	75	1 700
Dibenzofuran, Dibenzothiophene, Phenoxanthine, Acenaphthylene, 9-Fluorenone	135	3 320
Biphenyl, Diphenyl ether	209	9 179
Anthracene, Pyrene	287	15 065
Acenaphthene	299	13 202
Fluorene, 9H-Xanthene, 9H-Thioxanthene	407	19 925
Phenanthrene, Fluoranthene	527	29 645
9,10-Dihydro-anthracene	635	60 020
Chrysene, Benzo[k]fluoranthene	2 079	266 084
Benz[a]anthracene, Benzo[b]fluoranthene, Benzo[a]pyrene	4 095	531 440
Dibenzo[a,h]anthracene	8 255	3 392 577

with more advanced of *ab initio* methods that extend beyond the single-electron approximation (such as e.g. MP2, Møller-Plesset perturbation method including the second-order corrections), the calculation following the DFT system are considerably less time-consuming and require less computing resources (Puzyn & Falandysz, 2007; Puzyn, 2007). In spite of this fact, however, calculation of descriptors for congener families (very large, occasionally) at the DFT level still seems to be poorly effective. For example, calculation of molecular descriptors with B3LYP functional in 6-311++G(d,p) basis set for a relatively small group of 75 PCNs would require several days of work on seven advanced processors (Puzyn, 2009c). Hence, search for another more effective method was required.

In 2008, Puzyn proved that the most advanced semi-empirical methods (RM1 and PM6) were sufficient for QSPR modelling whereas the time needed for calculation of even very large groups of compounds would be reduced down to several hours (Puzyn, 2008). The proof relied on the re-calibration of previously published ten QSPR models in which the descriptors calculated at the DFT level were replaced one-by-one by the same descriptors calculated by means of semi-empirical RM1 and PM6 methods. Comparison of validation parameters of original models with the parameters of the models using the descriptors from semi-empirical calculations did not reveal any essential differences (Puzyn, 2008). It should be mentioned here that the descriptors present in original models were calculated using the B3LYP functional in basis sets featuring various sizes: 6-31G(d), 6-311G(d,p), and 6-311++G(d,p).

Practical methodology, which takes into account:

1. Combinatorial generation of internal coordinate matrices;
2. Calculation of descriptors at the level of semi-empirical methods; and
3. Determination of structural similarities of compounds by means of chemometric methods (Puzyn, 2008) considerably facilitated the process of modelling large sets (families) of chemical compounds having similar structures.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

QSPR-Based Model to Predict Subcooled Liquid (P_L) Vapour Pressure for POPs

Subcooled liquid (P_L) vapour pressure is a physical/chemical parameter of essential meaning from the point of view of Persistent Organic Pollutants transport processes in the natural environment via the atmospheric pathways. P_L values obtained through empirical studies are available for a low number of POPs congeners only (mainly Cl-POPs). Given a high number of possible congeners, effective determination of vapour pressure for each of them by way of empirical studies seems to be impracticable. As mentioned hereinabove, Br-POPs have an increasing impact on the natural environment; consequently, it appeared that the mobility of such compounds in the natural environment should be studied. Use of QSPR model is an effective alternative solution in experimental research works.

Gajewicz (2010) developed a QSPR model and used it for predicting the values of log P_L for 1436 polychlorinated congeners of polychlorinated and polybrominated POPs classified into 12 families (CBzs, BBzs, PCBs, PBBs, PCDDs, PBDDs, PCDFs, PBDFs, PCDEs, PBDEs, PCNs, and PBNs). Experimental values of P_L were available for 341 compounds, out of which 256 compounds were selected for the training set whereas 85 compounds made up the validation set. The modelling process was conducted by means of GA-PLS method, a modification of the PLS (Partial Least Squares) method, where selection of model variables (descriptors) was aided by the use of a genetic algorithm (Gramatica, 2010). Obtained model used four latent variables explaining in total 97% variability in vector **y** and 99% variability in matrix **X** (Gajewicz, 2010).

The analysis of loadings brought about by individual descriptors (8 descriptors selected by GA out of the pool of total 26 ones) to each of the latent vector allowed interpreting the model in physical terms. The number and the type of substituents were structural features that mostly determined the differences in vapour pressure for the modelled compounds. Vapour pressure decreases along with increase in the number of substituents whereas bromine derivative compounds feature smaller volatility compared with their chlorine derivative analogues. In the case of compounds having the same number of substituents of the same type, small differences in vapour pressure result from the differences in their positions. Vapour pressure increases along with increase in the number of substituents in *ortho* position, for example, vapour pressure of 2,2'-dibromodiphenyloether exceeds the vapour pressure of 4,4'-dibromodiphenyloether. Furthermore, molecule surface and volume-connected descriptors have a small meaning, too. The larger the surface and the volume, the smaller is the volatility of the molecule (Gajewicz, 2010).

The fact that all the 1436 compounds prove to have vapour pressure under 1000 Pa, that is related to high Long Range Transport Potential (LRTP) according to the Aarhus Protocol, seems an interesting conclusion drawn from the said studies.

QSPR-Based Model to Predict the Partition Coefficients (log K_{OW}, log K_{OA}, log K_{AW}) for POPs

The *n*-octanol/water (log K_{OW}), *n*-octanol/air (log K_{OA}), and air/water (log K_{AW}) partition coefficients are additional physical/chemical important quantities, which describe the mobility of chemical pollutants in the natural environment (Mackay, 2007, Mostrag, 2010; Puzyn & Falandysz, 2005; Puzyn, 2006). The values of the two first partition coefficients for 1436 polychlorinated or polybrominated POPs congeners were determined on the basis of QSPR modelling (Equations 8 and 9) (Puzyn, 2008b). The values of the first coefficient (log K_{AW}) were computed from the relationship expressed by the Equation 10.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

$$\begin{aligned}&\log K_{OW} = -0.3587(0.1521) - 0.1220(0.0238)\mu + 0.0247(0.0006)SAS\\&n = 178; k = 59; F = 1013; p < 0.001; R^2 = 0.920; Q^2_{CV} = 0.918; Q^2_{Ext} = 0.924\\&RMCEC = 0.315; RMSECV = 0.320; RMSEP = 0.302\end{aligned} \tag{8}$$

$$\begin{aligned}&\log K_{OA} = 7.3108(1.4538) + 0.7408(0.1554)\varepsilon_{HOMO} + 0.2862(0.0056)\alpha\\&n = 77; k = 26; F = 1287; p < 0.001; R^2 = 0.972; Q^2_{CV} = 0.970; Q^2_{Ext} = 0.961\\&RMCEC = 0.320; RMSECV = 0.333; RMSEP = 0.376\end{aligned} \tag{9}$$

$$\log K_{AW} = \log\left[\frac{K_{OW}}{K_{OA}}\right] = \log K_{OW} - \log K_{OA}. \tag{10}$$

The two models (Equations 10 and 11) were obtained by means of multiple regression method and featured a fine goodness-of-fit, robustness, and predictivity (Puzyn, 2008b).

The models used four quantum-mechanical descriptors calculated at the level of semi-empirical PM6 method, namely: dipole moment (μ); solvent accessible area (*SAS*); energy of the highest occupied molecular orbital (ε_{HOMO}), and average polarisation (α). Increased polarity of the compound (high value of dipole moment) decreases the value of logarithm of *n*-octanol/water (Equation 8), whereas increased *SAS* area related to the presence of high volumes of halogen substituents increases the value of log K_{OW}. Equation 9 contains the term ε_{HOMO} that in this case, describes the substitution pattern in the molecule, as well as the term α, a quantity that strongly depends on the number of substituents. Thus, the values of log K_{OA} coefficient increase along with increasing number of halogen substituents, whereas in the case of congeners having the same number of substituents, the differences are determined by the position of such substituents. For example, PCDDs congeners containing the substituents in positions 2, 3, 7, and 8 feature higher values of log K_{OA} compared with congeners substituted in positions 1, 4, 6, and 9 (Puzyn, 2008b).

Certain regularities were noted during the pairwise comparison of the predicted partition coefficients for chlorinated and brominated analogues (e.g. 1,2,3-trichlorodibenzo-*p*-dioxins and 1,2,3-tribromodibenzo-*p*-dioxins). This lead to simple method for of quick calculation of a given partition coefficient for brominated analogue (log K_{ij} [*Br*]), when the value of this coefficient (log K_{ij} [*Cl*]) is known for its chlorinated analogue (Equations 11-13) (Puzyn, 2008b):

$$\log K_{OW}\left[Br\right] = \log K_{OW}\left[Cl\right] + 0.44n_{Cl\to Br} \tag{11}$$

$$\log K_{OA}\left[Br\right] = \log K_{OA}\left[Cl\right] + 0.32n_{Cl\to Br} \tag{12}$$

$$\log K_{AW}\left[Br\right] = \log K_{AW}\left[Cl\right] - 0.21n_{Cl\to Br}. \tag{13}$$

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Although the quick method is obviously not as accurate as the QSPR modelling, it provides a first approximation of the coefficient value with no need to make quantum-mechanical calculations required to obtain the structural descriptors (Puzyn, 2008b).

Environmental Mobility of POPs Chlorine and Bromine Derivatives Assuming their Perfect Persistence

The estimation of environmental mobility of POPs is the next step in the process of risk assessment. Studies on mobility of chlorine and bromine POPs derivatives began from the similarity analysis of predicted partition coefficients following the averaging within the congener groups having the same number of substituents (e.g. mono-CDDs, di-CDDs, tri-CDDs, etc.) (Puzyn, 2008b). The similarity analysis was conducted with the two-way Hierarchical Cluster Analysis (two-way HCA), in which the Ward method was applied clustering, whereas Euclidean distance was used as the measure of similarity. The analysis produced five distinctive POPs classes featuring similar profile of distribution among the individual environmental components (Table 3).

Environmental mobility of chemical pollutants can be also studied with Multimedia Mass-balance (MM) models. MM models are research tools ensuring the studies on the impact of individual environmental components on the transport and fate of chemical substances (Hollander, 2008; Fenner, 2005; OCED, 2002, 2004). Furthermore, they allow grouping environmental pollutants following their potential for long range transport and persistence (Klasmeier, 2006). Consequently, they are valuable sources of information required for comprehensive risk assessment of new chemical substances. The terms "multimedia" or "multi-component" models explain their structure: the environment is divided into smaller elements (components) among which the substances are transported. The construction of MMs is based on setting-up of mass-balance equations for a set of interconnected 'boxes' that represent particular environmental compartments (i.e., surface water, deep oceanic waters, atmosphere etc.). Concentrations of chemicals in particular media and relevant rate constants (i.e., kinetic and transport parameters) are

Table 3. Relationship between the number of halogen atoms and membership in the similarity classes for the studied group of compounds (Puzyn, 2008b)

Group	Class I	Class II	Class III	Class IV	Class V
PCDD		0 – 1	2 – 5	6 – 8	
PBDD		0 – 1	2 – 3	4 – 5	6 – 8
PCDF		0 – 2	3 – 6	7 – 8	
PBDF		0 – 1	2 – 4	5 – 6	7 – 8
PCDE		0 – 1	2 – 6	7 – 10	
PBDE		0 – 1	2 – 4	5 – 6	7 – 10
PCB		0 – 2	3 – 7	8 – 10	
PBB		0 – 1	2 – 4	5 – 7	8 – 10
PCN	0	1 – 4	5 – 7	8	
PBN	0	1 – 3	4 – 5	6 – 7	8
CBz	0 – 4	5 – 6			
BBz	0 – 3	4 – 5	6		

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

used for generating the mass-balance equations. Alternative methodology is based on the concentration surrogate, namely fugacity (Mackay, 1979), which can be defined for a chemical as its 'tendency to escape' from the environmental phase. The fugacity is expressed in the units of pressure (it can be considered as a partial pressure proportional to the concentration). Moreover, the model includes emission and degradation processes inside each of the components. The complexity of MM models depends on thermodynamical assumptions to be applied and regional scale. There are four distinctive levels of model complexity according to the thermodynamical assumptions:

Level I: Closed system, state of equilibrium, steady state;
Level II: Open system, state of equilibrium, steady state;
Level III: Open system, no equilibrium, steady state; and
Level IV: Open system, no equilibrium, dynamical model. (OECD, 2002, 2004).

MMs can be also categorized into three main classes, dependently on, how much the simulated environment is regional-specific:

- **Generic (Standard) Models:** Models using default (average) parameters that characterise the environment, applied mostly in the fast screening environmental fate of new chemicals. They allow classifying the compounds into the groups of defined persistence and mobility in the environment, etc. Example: The OECD Tool (Wegmann, 2009, OCED, 2002; OECD, 2004).
- **Region-Specific Models:** Models using data that describe a defined area and are much more realistic than generic models. Examples: ELPOS (Beyer & Matthies, 2002), SimpleBOX (van der Meent, 1993), EUSES (Vermeire, 1997), CalTox (McKone, 1993), ChemCAN (Mackay, 1996).
- **Multi-Zone Models:** Models featuring more complex structure, allowing for classification into "sub-models" which represent individual areas or climatic zones. Models that consider: meteorological, hydrological, and geographical conditions, the specificities of emission patterns and population density in order to address the most realistic variability of the exposure distribution. Examples: GloboPOP (Wania & Mackay, 1995), CliMoChem (Scheringer, 2000), BETR (MacLeod, 2001, Fenner, 2005).

The input data required for developing MM models are as follow:

1. **Data Describing the Environment:** E.g. number and types of environmental components; "thickness" of individual components; average air and water temperature; average wind velocity, etc.;
2. **Data on the Emissions:** Volume or velocity of compound emissions; component exposed to emission; and
3. **Physical/Chemical Properties of the Modelled Compounds:** Most often: partition coefficients between air, water, soil, and half-life times in the particular compartments (OECD, 2002, 2004; Wegmann, 2009).

For the purpose of evaluating environmental mobility of the five POPs classes (previously identified with two-way HCA) a partitioning map was applied. The map had been originally developed by Wania (2006) on the basis of modelling the hypothetical chemical compounds with the GloboPOP model. It was assumed that the studied compounds would not be exposed to the process of degradation (perfect

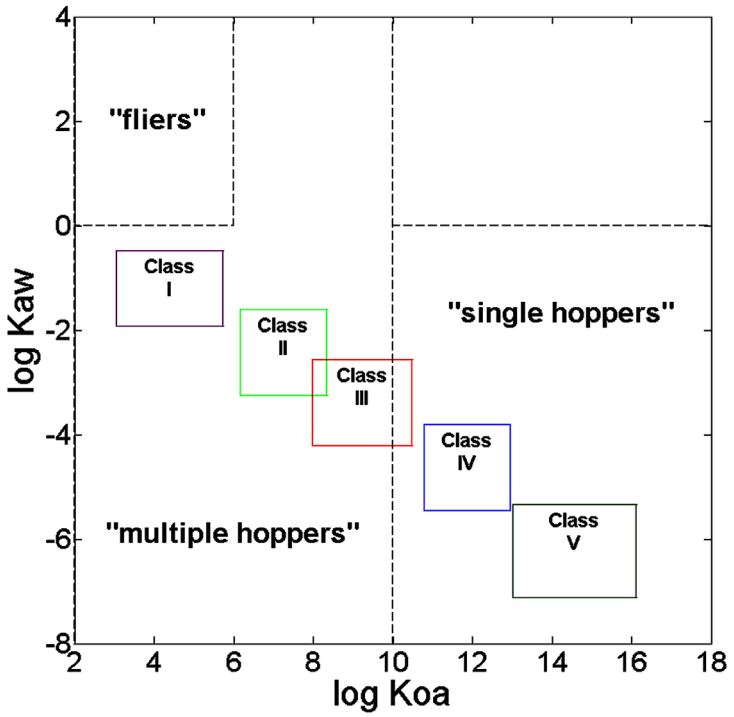
EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

persistence assumption). Thus, the only properties that influence the environmental mobility of the chemicals were the partitioning coefficients (degradation half-life times in the particular compartments were not taken into account in this case). Globo-POP is a level IV multi-zone model developed by the team of Frank Wania at the University of Toronto Scarborough (Wania & Mackay, 1995; Wania & Daly, 2002; Wania, 2006). The need to assume the perfect persistence resulted from shortage of data on the degradation (air, water, soil, and deposit half-life times). With the said assumption, the results of modelling lead to overestimation of mobility of compounds among the individual components and hence, make up a simulation according to the "worst-case scenario".

Environmental mobility of POPs decreases from Class I to Class V (Figure 3) (Puzyn, 2008b). Interestingly, the differences between the mobility values for chlorinated and brominated analogues (e.g. PCDDs and PBDDs) become specifically essential when comparing the compounds having a considerable number of substituents (e.g. hepta-CDDs belong to Class IV whereas hepta-BDDs are eligible for Class V) (Puzyn, 2008b).

Class I and II and partly, Class III compounds may move to and from the atmosphere, hydrosphere, and lithosphere through evaporation and next, through dry or wet deposition. In that way, they may be transported, inclusive of air masses, to the regions located far away from the sources of emission, towards higher or lower climatic zones. This phenomenon is known as the "grasshopper effect" (Gouin, 2004). On the other hand, Class IV and V compounds, because or their low mobility, will be quickly deposited from the air to the soil very close to the source of emission.

Figure 3. Partitioning map for the studied compounds fitted categories of environmental mobility (class I – class V)
(With copyright permission from Puzyn, 2008b).



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

In that way, the environmental mobility of 1436 POPs chlorine and bromine derivatives was simultaneously evaluated for the first time. The mobility evaluation was conducted on the basis of the proposed QSPR-MM hybrid modelling pattern. Accordingly, physical and chemical data that describe a given substance are determined by means of QSPR methods. Next, data are used in the MM model. If the individual partition coefficients were determined experimentally (not predicted with QSPR), investigations of environmental mobility of such high a number of substances would prove impossible in a reasonable period of time. This is the main advantage of using QSPR for supporting MM studies.

Predicting POPs Persistence in Environmental Components

The studies presented in previous chapter relied on the assumption of perfect persistence of modelled compounds. Although this assumption ensures the identification of the risk resulting from the process of transport and dissemination of POPs, it may, however, lead to excessive pessimistic forecasts since in the real world such compounds are exposed to environmental degradation processes.

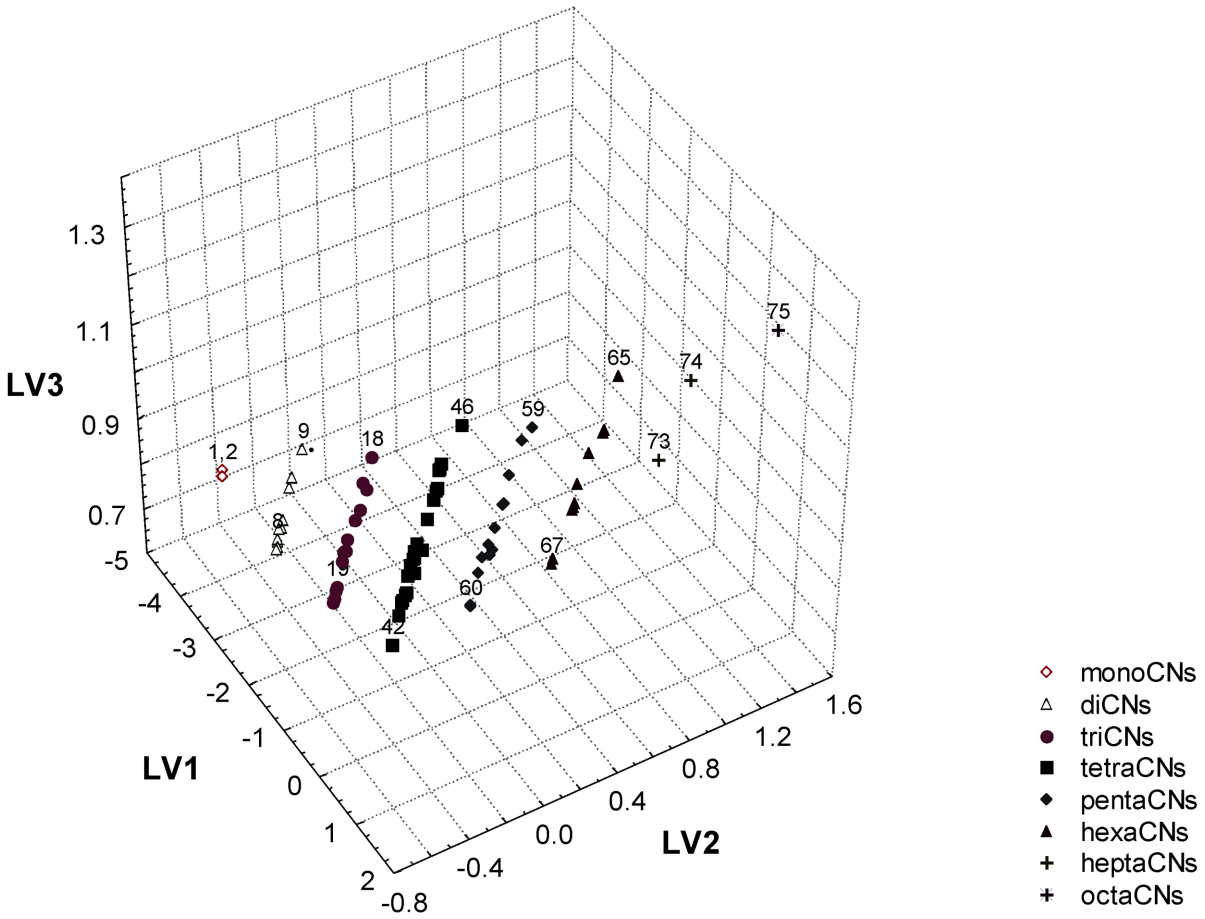
The atmosphere is the most important and effective POPs "transport corridor" in the natural environment. In the daytime, atmospheric degradation of organic compounds occurs mainly through the reactions with hydroxyl radicals. Concentration of hydroxyl radicals in the troposphere is relatively high $\left[OH^{\cdot}\right]$ = 9.4 x 10^5 radicals / cm^{-3} (Prinn, 2001). Assuming that the $\left[OH^{\cdot}\right]$ concentration is not significantly changing in time, POPs degradation will occur in compliance with pseudo-first-order kinetics. Thus, the half-life time $t_{1/2}^{\left[OH^{\cdot}\right]}$ may be used as the quantity describing the persistence of a given compound (Mackay, 2007).

In 2008, Puzyn (2008) developed a QSPR model that predicts the values of $t_{1/2}^{\left[OH^{\cdot}\right]}$ for POPs (Puzyn, 2008c). The model, developed with GA-PLS method, was used for predicting the half-life times for PCNs, one of the priority groups from among seven new POPs families included in the Aarhus Protocol. Predicted values of $t_{1/2}^{\left[OH^{\cdot}\right]}$ ranged from 2 through 343 days. Persistence of the compounds increased due to the increasing number of substituents (chlorine atoms) in the molecule. It should be mentioned here that the values of $t_{1/2}^{\left[OH^{\cdot}\right]}$ determined for the majority of congeners (apart from mono-CNs) allow also classifying the compounds as "persistent" on the basis of one of the criteria defined in Aarhus Protocol and Stockholm Convention (air $t_{1/2} > 2$ days). Given the model predicting error (RMSEP = 0.14), we cannot exclude the case where the least persistent mono-CNs will also meet the said criterion.

A structural fragment typical for the least persistent compounds was identified among the congeners featuring the same number of substituents, thanks to the analysis of arrangement of PCNs congeners in the space defined by three hidden variables of GA-PLS model (Figure 4) as well as to the development of electrostatic potential maps (Figure 5). In this case, the summary reaction rate is determined by the stage of hydroxyl radical attack against the molecule. Carbon atom α will be specifically vulnerable to such attacks if the two hydrogen atoms bound to the neighbouring carbon atoms (α and β) are substituted by chlorine atoms (Figure 5). In this situation, the induction effects which acts in opposite directions will produce a partial positive charge on the carbon atom α and hence, facilitate the attack by $\left[OH^{\cdot}\right]$.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

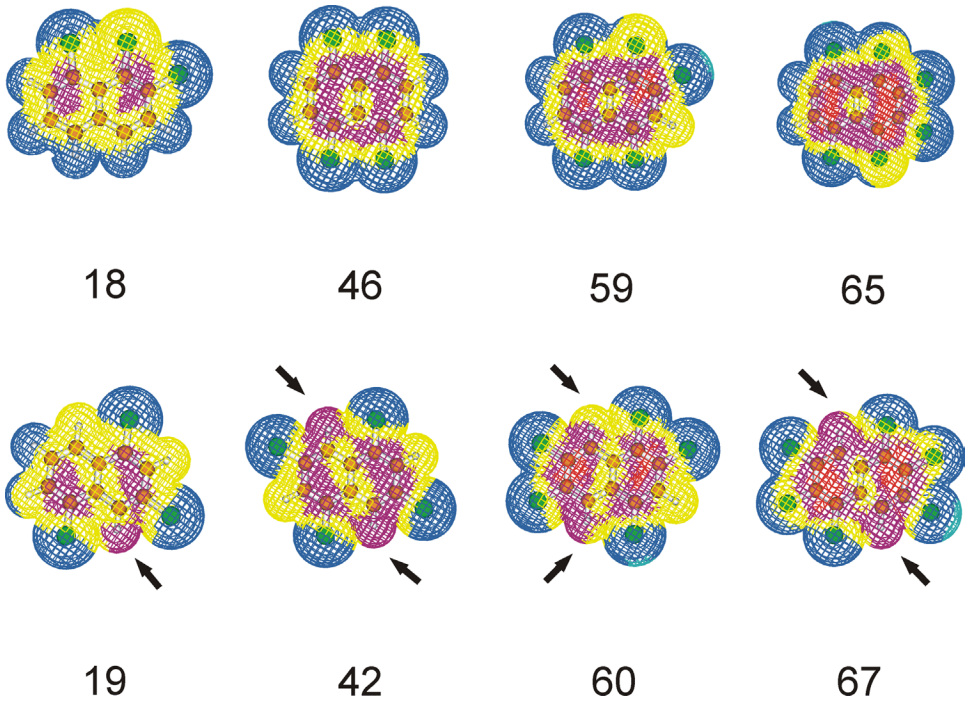
Figure 4. Score plot illustrating distribution of the individual congeners along with the latent vectors; the most and the least persistent congeners from each group are marked by the numbers.
(With copyright permission from Puzyn, 2008c).



The results concern the quantitative predictions of persistence of group POPs compounds on the basis of one of many particular degradation processes (i.e. reaction with hydroxyl radicals) occurring in the environment. In this case, QSPR modelling was feasible thanks to the availability of relatively high number of experimental data (experimental values of $t_{1/2}^{[OH\cdot]}$ for POPs). In practice, the procedure relying on the estimation of the "overall half-life time" for a defined environmental component (e.g. air $t_{1/2}$; water $t_{1/2}$, etc.) will be the most common approach. It takes into account all the most important chemical and biological processes that are responsible for compound degradation in a defined environmental component. The method of estimating the half-life time was proposed for the first time by Mackay (2007), who specified the air, water, soil, and sediments $t_{1/2}$ values for a dozen compounds (mainly chlorine derivatives) from POPs group in his handbook. The fact that a defined compound is assigned to a defined category allows assuming that its persistence will be equal to the average value of half-life time in a defined category, subject to a considerable approximation.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

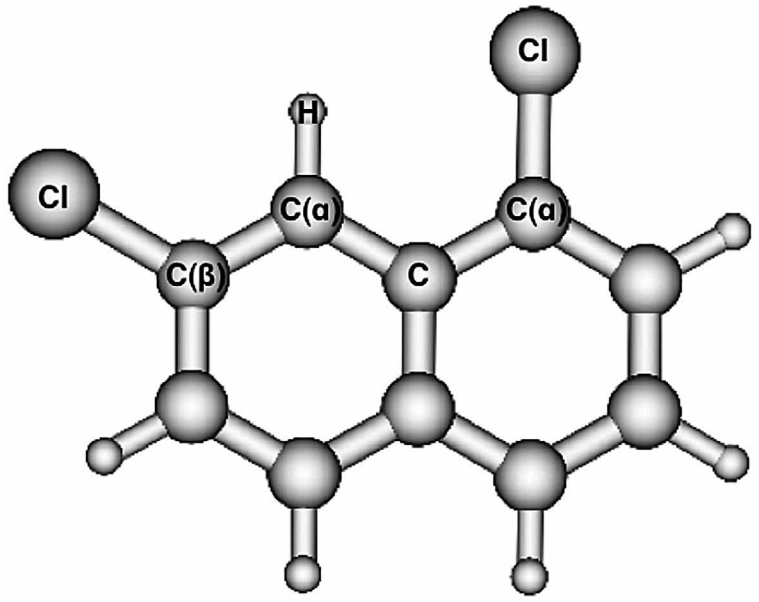
Figure 5. Maps of the electrostatic potentials along the most persistent (upper row) and the least persistent (lower row) representatives of tri-, tetra-, penta-, and hexa-CNs; a three-layer structure, in which the positively charged layers (dark red color) are ''closed'' in the central part of the molecules by surrounding neutral (light yellow color) and then negatively charged (dark blue color) external layers. Arrows mark a-hydrogen atoms with high positive partial charges (deficit of electrons) – the structural fragments where the ''insulating'' neutral layers ''open.''
(With copyright permission from Puzyn, 2008c).



Air, water, soil, and sediments half-life time values are available for most POPs chlorine derivatives. Persistence of bromine derivatives may be predicted on the basis of computational methods. Puzyn (2011) built four classification models for 1431 congeners of POPs bromine and chlorine derivatives as well as for unsubstituted polycyclic aromatic hydrocarbons; consequently, the models allowed assigning the individual congeners to "half-quantitative" persistence classes in the air, superficial water layers, as well as superficial soil and sediments (Puzyn, 2011). The process of modelling was conducted by means of the kNN method (k-Nearest Neighbours) coupled with the PCA (Principal Component Analysis). The most essential feature of this method is the application of principal components (from the PCA) as structural descriptors in kNN classification models. Linear combinations of 24 original topological and quantum-mechanical descriptors calculated at the level of semi-empirical PM6 method were used as the principal components. Classification accuracy was checked by means of detailed validation procedure with account for internal validation relying on multiple divisions into the training and (internal) validation sets as well as for external validation. The classification accuracy determined on the basis of external validation was, respectively, 86%, 85%, 87%, and 75% for the models describing air, superficial water, soil, and sediments persistence of compounds (Puzyn, 2011).

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 6. The structural –C(b)Cl–C(a)H–C–C(a)Cl– fragment
(With copyright permission from Puzyn, 2008c).



The above-mentioned study was the first attempt to estimate simultaneously POPs chlorine and bromine derivatives persistence in the individual environmental components (Puzyn, 2011). Each of the compounds was assigned to one of the five persistence classes: air persistence (A1-A5); water persistence (W1-W5), soil persistence (So1-So5), and to one of three sediment persistence classes (Se1-Se3) which, given the availability of calibration data, were minor modifications of ten classes proposed by Mackay in his handbook (Mackay, 2007) (Table 4).

Table 4. Both semi-quantitative classification systems proposed by Mackay (2007) and Puzyn (2011)

Mackay's Class (2007)			**Puzyn's Class (2011)**			
Range of t1/2 [h]	**Mean t1/2 [h]**	**Symbol**	**Air**	**Water**	**Soil**	**Sedeminets**
< 10		1		W1		
10-30	17	2	A1			
30-100	55	3				
100-300	170	4	A2	W2	So1	
300-1000	550	5	A3			
1000-3000	1700	6	A4	W3	So2	Se1
3000-10000	5500	7	A5	W4	So3	
10000-30000	17000	8			So4	Se2
30000-100000	55000	9		W5	So5	Se3
> 100000		10				

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

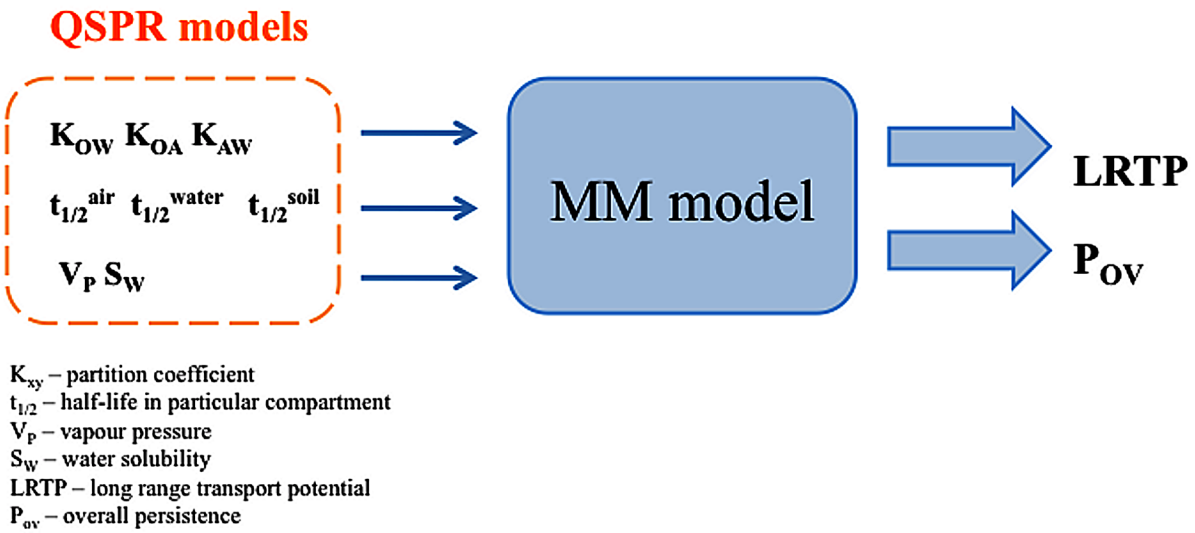
Persistence of chloroorganic compounds increased along with increasing number of chlorine atoms in the molecule. On the other hand, in the case of brominated compounds, the same tendency was observed in the air and the deposits. Considering high photo-reactivity of C-Br bonds, the processes of photolytic degradation of POPs bromine derivatives in superficial water layers may disturb the tendency in these two components. The availability of data on persistence (half-life time values) in the individual environmental components and of the values of partition coefficients for POPs makes feasible a more realistic modelling the processes of environmental transport and deposition of such compounds on the basis of MM models (departure outside the "perfect persistence approximation").

Comparison of QSPR-MM Modelling Scheme with MM Modelling Using Experimental Data

Input data for a MM model that describes physical/chemical properties of compounds may be either determined experimentally or taken from QSPR modelling. The OECD Guidelines (OECD, 2002, 2004) prefers using experimental data. QSPR-MM modelling scheme (Figure 7) in which the experimental data are replaced by data from QSPR proved no essential degeneration of the results of MM modelling (Puzyn, 2011).

In this work Puzyn (2011) used *The Tool 2.2* – a simple mass-balance multimedia model recommended by OECD in order to predict the LRTP (long-range transport potential) and P_{OV} (overall persistence) of chemical compounds (Wegmann, 2009). *The Tool* is a level III standard model where the environment is divided into three compartments: soil, water and troposphere. Each of the compartments is composed of smaller components: soil contains solid particles, pore water, and pore-filling air; water contains pelagic zone and suspended organic matters; troposphere contains air and aerosol particles. For the purpose of environmental description, the most important quantities are:

Figure 7. Scheme of QSPR-MM model development



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

1. Assumption that 71% of surface is covered with waters;
2. Atmospheric component height amounts to 6 km;
3. Superficial soil layer depth amounts to 100 cm;
4. Superficial water layer depth amounts to 100 m;
5. Wind velocity amounts to 4 m s^{-1}; and
6. Water flow velocity amounts to 0.02 m s^{-1}.

The model supplies information on the total time of compound staying in the environment, i.e. the total persistence P_{OV} (in days), and the LRTP expressed in two ways: as CTD, i.e. Characteristic Travel Distance, and as TE, i.e. Transfer Efficiency. The first measure of LRTP, i.e. *CTD*, is defined as the distance (in km) from the source of emission where the substance level decreases down to 1/e (approximately 37%). The second measure of LRTP, i.e. *TE*, expresses (in per cent) such part of substance emission that will arrive at a defined point (in this case, at the second part of the model environment if the latter is symmetrically divided into two parts composed of the same components) (Wegmann, 2009).

In Table 5 there are summarized the calculations of P_{OV}, *CTD*, and *TE* by using six MM models that relied on input data from various sources. The MM model where both the partition coefficients and the values of half-life times were determined experimentally (model I) was used as a reference model. Models II and V employed the partition coefficients predicted by QSPR models (Puzyn, 2008b), whereas models IV-III used data on persistence from classification models (Puzyn, 2011). Additionally, two models (III and VI) applied partition coefficients predicted by *EPISuite* software package developed by the US Environmental Protection Agency (EPISuite, 2010). *EPI Suite* is a tool widely applied for the risk assessment. Although *EPISuite* relies on simple QSPR models, the application is a specific "black box" and the user is unable to evaluate the quality of results (for instance, to check whether, or not, a defined compound falls to the applicability domain).

Comparison of the values of P_{OV}, *CTD*, and *TE* determined by means of MM models utilizing input data from calculation, with the reference model (model I), was conducted by means of the Student's t-test for paired samples (in the following configurations: model I – model II; model I – model III; model I – model IV, etc.) and did not show statistically significant differences ($p < 0.05$). On the other hand, the analysis of ranks correlation in pairs (model I – model II; model I – model III, etc.) with use of Spearman's ρ and Kendall's τ statistics allowed detecting that even minor differences in the values of P_{OV}, *CTD*, and *TE* determined by means of calculation of individual models could not affect the order of compounds in the environmental persistence and/or mobility series (Figure 8). Specifically, replacement of experimental half-life times by those predicted by QSPR (Puzyn, 2011) had higher impact on the resulting total P_{OV} persistence series than the replacement of partition coefficients. It should be mentioned, however, that the differences concerned single compounds in the series and in the same way, could result from a considerable cross-similarity of POPs congeners in terms of their pathways in the environment (on the basis of the values of P_{OV}, *CTD*, and *TE*).

All the above-mentioned observations lead to a conclusion that the differences between the results of MM modelling that employs experimental data and of the QSPR-MM pattern are of minor importance. When considering the ability to perform modelling the environmental mobility of a high number of compounds by means of QSPR-MM methods at the same time, this method may be recommended as a sound substitute for the existing MM modelling pattern that relies on experimental data only in compliance with OECD recommendations.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 8. Rank correlations between the outputs from multimedia models based on the predicted physical/chemical data (II–VI) and the outputs from the model utilizing only the experimental properties (I) *(With copyright permission from Puzyn, 2011).*

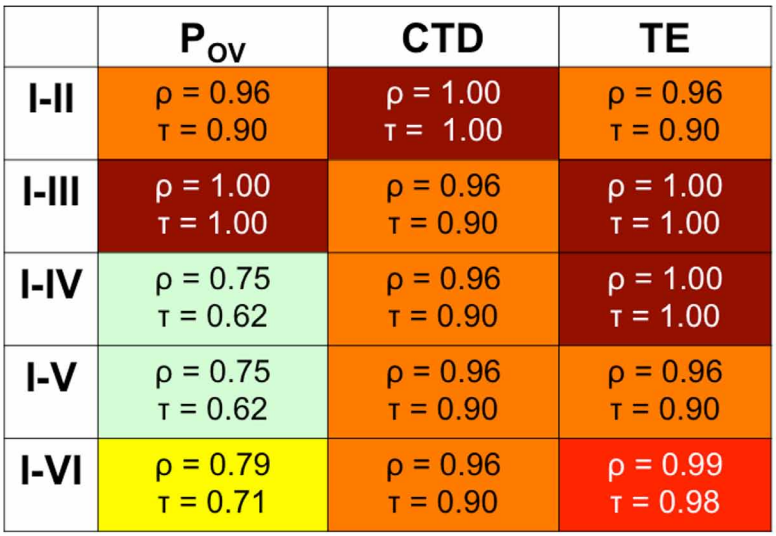


Table 5. Compared MM models (Puzyn, 2011)

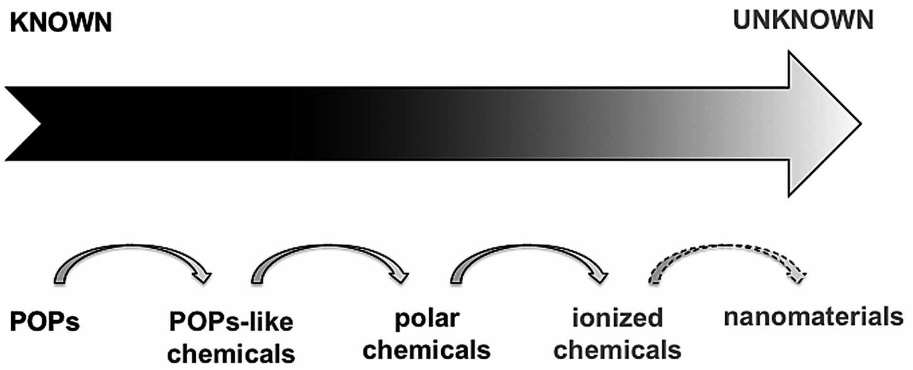
Model	Partition Coefficient	Half-Time Life
I	Experimentally measured	Experimentally measured
II	Predicted by QSPR (Puzyn, 2008b)	Experimentally measured
III	Predicted by EPI Suite	Experimentally measured
IV	Experimentally measured	Predicted by QSPR (Puzyn, 2011)
V	Predicted by QSPR (Puzyn, 2008b)	Predicted by QSPR (Puzyn, 2011)
VI	Predicted by EPI Suite	Predicted by QSPR (Puzyn, 2011)

PERSPECTIVE OF FURTHER DEVELOPEMENT OF COMPUTATIONAL TECHNIQUES FOR ENVIRONMENTAL EXPOSURE ASSESSMENT

The main disadvantage of the Multimedia Mass-balance models developed so far is that, they are applicable for typically hydrophobic and non-ionized compounds (i.e. POPs and POPs-like chemicals). However, regarding the widespread detection of polar and ionized compounds, such as: Pharmaceuticals and Personal Care Products (PPCPs) in various environmental compartments (Carlsson, 2006a, b; Cooper, 2008; Daughton & Ruhoy, 2009, 2011; Halling-Sorensen, 1998), the further efforts should be taken to extend applicability of QSPR-MM to model environmental fate of those chemicals. Moreover, considering the rapid increase in the production and usage of nanomaterials (NMs) – it would be highly beneficial to develop QSPR-MM modelling scheme also for nanoparticles constituting NMs as a new and important group of environmental pollutants (Gottschalk, 2013; Nowack, 2012; Sun, 2014). The conceptual framework of further developing QSPR-MM for environment exposure assessment is presented in Figure 9.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 9. Scheme of the conceptual framework of computational techniques application in environmental exposure assessment



Some partial adjustments have been implemented in the existing MMs to expand its usage for more polar and/or ionized compounds. The modelling results performed so far (Zukowska 2006) indicate that polar and ionized chemicals are rather single-media pollutants, prone to remain in the compartment of emission (usually in water or soil). Thus, for such compounds, single-media approaches can be applied to model their environmental transport and fate. According to EU Scientific Committee on Toxicity, Ecotoxicity and the Environment (CSTEE, 2001), two single-media models are especially recommended: i) The European Union System for the Evaluation of Substances, EUSES (Vermeire 1997) and (ii) The Geo-referenced Regional Exposure Assessment Tool for European Rivers, GREAT-ER (Schowanek, 2001; Kormann, 2006). Those models have already been applied to study environmental fate of selected pharmaceuticals (Wezel & Jager, 2002; Schowanek & Webb, 2002).

First, conclusions that might be drawn based on those studies are as follows. First, it is necessary to make some corrections in the models developed so far in order to include much more comprehensive description of environmental transport phenomena occurring in case of highly polar and/or ionized compounds. This means the selection of input physical/chemical parameters used in MM models so far should be reconsidered. For example, the partition coefficients of ionic compounds should be used only for the fraction of undissociated species. In this case, pK_a and pH should be also included.

Second, it is still difficult to estimate the environmental mass-balance of PPCPs in a realistic way, since pharmaceuticals are easily metabolised in the human body and by other living organisms; this process should not be negligible.

Third, the possible additional source of the modelling uncertainty may be associated with degradation and loss processes occurring much faster for most PPCPs than for POPs and POPs-like unpolar compounds. These are for example: the removal during sewage treatment, photodegradation, sorption to sediments, etc. Availability of data describing these processes is frequently very limited. The limitation concern also experimentally determined values of physical/chemical properties (properties utilized as the inputs parameters in MM models) that are necessary to assess the environmental persistence and fate.

Also due to the enormous structural diversity of PPCP chemicals, the application of QSPR models can be much more problematic than the application of QSPR for POPs. Majority of the published QSPR models (Bergstrom, 2004; Cao, 2010; Chen, 2002; Du-Cuny, 2008; Duchowicz & Castro, 2009; Ghasemi & Saaidpour, 2007; Hansen, 2006) do not fulfil the OECD recommendations (see Section 2 of this chapter). Most often the existing models do not have appropriately designed applicability domain,

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

there is a lack of external validation or even lack of validation at all. The models were frequently developed for large, structurally diversified groups of compounds, without appropriate representation of the structural diversity in the training set. This problem is usually reflected in high errors of performed predictions (e.g. errors of predicted aqueous solubility with the model developed by Du-Cuny (2008) for 2473 drug-like compounds are up to 1.0 log unit).

Therefore, in the case of PPCPs, further, more comprehensive and systematic studies are necessary not only in the field of MM development, but also to improve the existing QSPR tools. In our opinion, special emphasis should be placed on the development of a set of simple local QSPR models that cover the whole domain of chemically diversified compounds (i.e. local models should be developed separately for particular chemical classes of compounds) rather than a single model that predicts given phys/chem property for the whole group of PPCPs.

As mentioned at the beginning of this section, besides PPCPs, there is a growing interest in modelling environmental fate of nanomaterials. Emission of these compounds into the environment may occur throughout their entire life cycle: from the production, application up to the end of its life (Sun, 2014). It is expected that such materials, deposed in significant levels into the environment, might create such as significant environmental problems as POPs in last decade (Gottschalk, 2013). Therefore, environmental exposure assessment for nanomaterials is highly required.

All studies dealing with environmental behaviour of nanomaterials published so far were reviewed by Gottschalk (2013). The authors indicate that due to insufficient amount of empirical data, influencing the environmental behaviour of NMs (partition coefficients, half-live times, vapour pressure, enthalpies of phase transfer, etc.), application of computational techniques to model fate and transport of nanomaterials is still very limited. Before starting the modelling, problem with data availability have to be solved. To do that, new analytical techniques for detecting, characterizing and quantifying nanomaterials in natural ecosystems must be developed first (Hassellov & Kaegi, 2009, Mitrano, 2012, von der Kammer, 2012). Considering fact that the existence of different physical forms and sizes of the same material influent their physical/chemical properties, it is not a trivial task. And this task should be performed with close collaboration between experimentalists and computational specialists. It is important because the modellers would not only analyse/use the data after completing the experiment, but also they could significantly contribute to the planning stage (Puzyn, 2009b). In this way, the obtained data would be much more useful for developing QSPR-MM models. Application of QSPR methodology for nanomaterials (Nano-QSPR) requires solving one additional problem: the appropriate characterization of nanoparticles' structure.

Although more than 5000 structural descriptors have been defined so far, they might be insufficient to describe the uniqueness of nanoparticles. In our opinion, novel and more appropriate types of the descriptors should be developed. Nanodescriptors should reflect not only molecular structure but also supra-molecular pattern (e.g. size, shape, porosity, etc.) of these materials and can be derived from both computational and experimental techniques (Puzyn, 2009b).

It should be emphasised, however, that despite the limitation, there are several Nano-QSPR models developed so far for nanomaterials. For example, Toropov (2007) developed two models that allow to predict two fundamental, from the environmental point of view, physical-chemical properties of carbon nanotubes, namely water solubility and *n*-octanol/water partition coefficient. These models have good statistics but might be unstable since the ratio of used descriptors (two-element chiral vectors) to compounds in training set (8 compounds) is low (Topliss ratio). Also, Martin (2007) proposed models to predict the solubility of buckminsterfullerene in *n*-heptane and in *n*-octanol. The applicability domain

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

of these models was not validated, but considering the structural diversity between compounds utilized to calibrate models (polycyclic aromatic hydrocarbons that are planar) and fullerenes (spherical) performed prediction may not be reliable. Thus, the main disadvantage of published models is that they do not fulfil all OECD criteria of model's quality (see the second section of this chapter). However, in our opinion, criteria developed for "bulk" chemicals may not be sufficient for chemicals in "nano" scale. This is because the number of systematic experimental data available for NMs obtained with the same protocol is usually too small to split them into training and validation set. Considering that, the novel, minimum requirements for validation of Nano-QSPR models have to be evaluated.

CONCLUSION

Recently, a novel modelling methodology that links QSPR and MM has been implemented. According to this approach, the physical/chemical properties required as the input variables for multimedia modelling can be calculated directly from appropriate QSPR models. QSPR-MM modelling scheme was successfully applied in exposure assessment of POPs. As it was demonstrated, there are no significant differences between the results of MM modelling that employed experimental data and those one predicted by QSPR models. Moreover, in this case, computational predictions might be significantly supported by the use of combinatorial chemistry tools like ConGENER software package.

In the case of more polar and/or ionized chemicals further efforts has to be taken to improve the prediction of P_{OV} and LRTP. This effort should not only include much more comprehensive description of environmental transport phenomena important for highly polar and/or ionized compounds, but also the special emphasise should be put in the set and sources of input parameters. Majority of QSPR models developed so far for PPCPs are not property validated. Thus, their results are disputable. They might be rather considered as a starting point for the further development of appropriate modelling methodology for this class of compounds.

Greater effort is required in the case of exposure assessment of nanomaterials. In this field, the knowledge about parameters fundamental for environmental fate and transport is still insufficient. Moreover, available techniques, originally developed for bulk chemicals, are not sufficient for compounds in "nano" scale. In case of this class of pollutants, further effort is necessary to improve and/or adopt experimental techniques in order to obtain data that could be employed in QSPR-MM methodology.

Stronger collaboration between experimentalists and modellers is recommended not only in the case of nanomaterials, but also for investigating transport and fate of other chemicals in order to reach compromise between the highest possible (from the economical point of view) number of compounds selected for experiment and the lowest number of compounds required to develop reliable QSPR models.

ACKNOWLEDGMENT

The authors are grateful for the financial support of the European Commission through Marie Curie IRSES program: NanoBRIDGES project (FP7-PEOPLE-2011-IRSES, Grant Agreement Number 295128) and NanoPUZZLES project (Grant Agreement Number 309837). This paper was prepared with financial support of the Foundation for Polish Science (FOCUS No F2/2010/P/2013) and the Polish Ministry of Science and Higher Education Grant No. DS 530-8180-D202-12. KJ thanks the European Social Fund

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

for granting her with a fellowship in frame of the project ‘‘Development Program of the University of Gdansk in areas of Europe 2020 (UG 2020)’’ supported by Human Capital Operational Programme, Action 4.3, Strengthening of didactic potential of universities in key areas in the context of the goals of the Europe 2020 strategy (Grant No. UDA – POKL. 04.03.00-00-047/12).

REFERENCES

Batterman, S., Chernyak, S., Gwynn, E., Cantonwine, D., Jia, C., Begnoche, L., & Hickey, J. P. (2007). Trends of brominated diphenyl ethers in fresh and archived Great Lakes fish (1979-2005). *Chemosphere*, *69*(3), 444–457. doi:10.1016/j.chemosphere.2007.04.066 PMID:17561231

Bergstrom, C. A. S., Wassvik, C. M., Norinder, U., Luthman, K., & Artursson, P. (2004). Global and local computational models for aqueous solubility prediction of drug-like molecules. *Journal of Chemical Information and Computer Sciences*, *44*(4), 1477–1488. doi:10.1021/ci049909h PMID:15272856

Beyer, A., & Matthies, M. (2002). *Environmental long-range transport and persistence of substances. ELPOS 1.0.1. institute for environmental systems research*. Germany: University of Osnabruck.

Birnbaum, L. S., Staskal, D. F., & Diliberto, J. J. (2003). Health effects of polybrominated dibenzo-p-dioxins (PBDDs) and dibenzofurans (PBDFs). *Environment International*, *29*(6), 855–860. doi:10.1016/S0160-4120(03)00106-5 PMID:12850101

Cao, D. S., Xu, Q. S., Liang, Y. Z., Chen, X., & Li, H. D. (2010). Prediction of aqueous solubility of druglike organic compounds using partial least squares, back-propagation network and support vector machine. *Journal of Chemometrics*, *24*, 584–595.

Carlsson, C., Johansson, A. K., Alvan, G., Bergman, K., & Kuhler, T. (2006a). Are pharmaceuticals potent environmental pollutants? Part I: Environmental risk assessments of selected active pharmaceutical ingredients. *The Science of the Total Environment*, *364*(1-3), 67–87. doi:10.1016/j.scitotenv.2005.06.035 PMID:16257037

Carlsson, C., Johansson, A. K., Alvan, G., Bergman, K., & Kuhler, T. (2006b). Are pharmaceuticals potent environmental pollutants? Part II: Environmental risk assessments of selected pharmaceutical excipients. *The Science of the Total Environment*, *364*(1-3), 88–95. doi:10.1016/j.scitotenv.2005.06.036 PMID:16260028

CEPA. (1999). *The Canadian environmental protection act the government of Canada*. Canada: CEPA.

Chen, X.-Q., Cho, S. J., Li, Y., & Venkatesh, S. (2002). Prediction of aqueous solubility of organic compounds using a quantitative structure–property relationship. *Journal of Pharmaceutical Sciences*, *91*(8), 1838–1852. doi:10.1002/jps.10178 PMID:12115811

Chirico, N., & Gramatica, P. (2012). Real external predictivity of QSAR models. Part 2. New intercomparable thresholds for different validation criteria and the need for scatter plot inspection. *Journal of Chemical Information and Modeling*, *52*(8), 2044–2058. doi:10.1021/ci300084j PMID:22721530

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Cooper, E. R., Siewicki, T. C., & Phillips, K. (2008). Preliminary risk assessment database and risk ranking of pharmaceuticals in the environment. *The Science of the Total Environment*, *398*(1-3), 26–33. doi:10.1016/j.scitotenv.2008.02.061 PMID:18440598

Cronin, M. T. D. (2010). Quantitative structure-activity relationships (QSAR) - Applications and methodology. In T. Puzyn, J. Leszczynski, & M. T. D. Cronin (Eds.), *Recent advances in QSAR studies. Challenges and advances in computational chemistry and physics*. Dordrecht, The Netherlands: Springer. doi:10.1007/978-1-4020-9783-6_1

Cronin, M. T. D., & Schultz, T. W. (2003). Pitfalls in QSAR. *Journal of Molecular Structure THEOCHEM*, *622*(1-2), 39–51. doi:10.1016/S0166-1280(02)00616-4

CSTEE. (2001). Environmental risk assesment of medicinal products for human use. Brussels, Belgium: Author.

Daughton, C. G., & Ruhoy, I. S. (2009). Environmental footprint of pharmaceuticals: The significance of factors beyond direct excretion to sewers. *Environmental Toxicology and Chemistry*, *28*(12), 2495–2521. doi:10.1897/08-382.1 PMID:19382823

Daughton, C. G., & Ruhoy, I. S. (2011). Green pharmacy and pharmEcovigilance: Prescribing and the planet. *Expert Review and Clinical Pharmacology*, *4*(2), 211–232. doi:10.1586/ecp.11.6 PMID:22115404

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

Du, B., Zheng, M., Huang, Y., Liu, A., Tian, H., Li, L., & Su, G. et al. (2010). Mixed polybrominated/chlorinated dibenzo-p-dioxins and dibenzofurans in stack gas emissions from industrial thermal processes. *Environmental Science & Technology*, *44*(15), 5818–5823. doi:10.1021/es100867d PMID:20586446

Du-Cuny, L., Huwyler, J., Wiese, M., & Kansy, M. (2008). Computational aqueous solubility prediction for drug-like compounds in congeneric series. *European Journal of Medicinal Chemistry*, *43*(3), 501–512. doi:10.1016/j.ejmech.2007.04.009 PMID:17574307

Duchowicz, P. R., & Castro, E. A. (2009). QSPR studies on aqueous solubilities of drug-like compounds. *International Journal of Molecular Sciences*, *10*(6), 2558–2577. doi:10.3390/ijms10062558 PMID:19582218

EC. (2006). *Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerming the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, and amemding Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC*. Brussels: The European Parliament and the Council of the European Union.

EPISuite. (2010). Estimation programs interface suite™ for Microsoft® Windows, v 4.00. United States Environmental Protection Agency.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Eriksson, L., Jaworska, J., Worth, A. P., Cronin, M. T. D., McDowell, R. M., & Gramatica, P. (2003). Methods for reliability and uncertainty assessment and applicability evaluations of classification- and regression-based QSARs. *Environmental Health Perspectives*, *111*(10), 1361–1375. doi:10.1289/ehp.5758 PMID:12896860

Falandysz, J. (1999). *Polichlorowane bifenyle (PCBs) w środowisku. chemia, analiza, toksyczność, stężenia i ocena ryzyka*. Gdańsk: Fundacja Rozwoju Uniwersytetu Gdańskiego.

Fenner, K., Scheringer, M., Macleod, M., Matthies, M., McKone, T., Stroebe, M., & Wania, F. et al. (2005). Comparing estimates of persistence and long-range transport potential among multimedia models. *Environmental Science & Technology*, *39*(7), 1932–1942. doi:10.1021/es048917b PMID:15871221

Gajewicz, A., Haranczyk, M., & Puzyn, T. (2010). Predicting logarithmic values of the subcooled liquid vapor pressure of halogenated persistent organic pollutants with QSPR. How different are chlorinated and brominated congeners? *Atmospheric Environment*, *44*(11), 1428–1436. doi:10.1016/j.atmosenv.2010.01.041

Ghasemi, J., & Saaidpour, S. (2007). QSPR prediction of aqueous solubility of drug-like organic compounds. *Chemical & Pharmaceutical Bulletin*, *55*(4), 669–674. doi:10.1248/cpb.55.669 PMID:17409570

Gobas, F. A., de Wolf, W., Burkhard, L. P., Verbruggen, E., & Plotzke, K. (2009). Revisiting bioaccumulation criteria for POPs and PBT assessments. *Integrated Environmental Assessment and Management*, *5*(4), 624–637. doi:10.1897/IEAM_2008-089.1 PMID:19552497

Gottschalk, F., Sun, T. Y., & Nowack, B. (2013). Environmental concentrations of engineered nanomaterials: Review of modelling and analytical studies. *Environmental Pollution*, *181*, 287–300. doi:10.1016/j.envpol.2013.06.003 PMID:23856352

Gouin, T., Mackay, D., Jones, K. C., Harner, T., & Meijer, S. N. (2004). Evidence for the "grasshopper" effect and fractionation during long-range atmospheric transport of organic contaminants. *Environmental Pollution*, *128*(1-2), 139–148. doi:10.1016/j.envpol.2003.08.025 PMID:14667724

Gramatica, P. (2007). Principles of QSAR models validation: Internal and external. *QSAR & Combinatorial Science*, *26*(5), 694–670. doi:10.1002/qsar.200610151

Gramatica, P. (2010). Chemometric methods and theoretical molecular descriptors in predictive QSAR modeling of the environmental behavior of organic pollutants. In T. Puzyn, J. Leszczynski, & M. T. D. Cronin (Eds.), *Recent advances in QSAR studies. Challenges and advances in computational chemistry and physics*. Dordrecht, The Netherlands: Springer. doi:10.1007/978-1-4020-9783-6_12

Halling-Sorensen, B., Nors Nielsen, S., Lanzky, P. F., Ingerslev, F., Holten Lutzhoft, H. C., & Jorgensen, S. E. (1998). Occurrence, fate and effects of pharmaceutical substances in the environment--a review. *Chemosphere*, *36*(2), 357–393. doi:10.1016/S0045-6535(97)00354-8 PMID:9569937

Hansen, N. T., Kouskoumvekaki, I., Jørgensen, F. S., Brunak, S., & Jonsdottir, S. O. (2006). Prediction of pH-dependent aqueous solubility of druglike molecules. *Journal of Chemical Information and Modeling*, *46*(6), 2601–2609. doi:10.1021/ci600292q PMID:17125200

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Haranczyk, M., Puzyn, T., & Sadowski, P. (2008). ConGENER - A tool for modeling of the congeneric sets of environmental pollutants. *QSAR & Combinatorial Science*, *27*(7), 826–833. doi:10.1002/qsar.200710149

Hassellov, R., & Kaegi, R. (2009). Analysis and characterization of manufactured nanoparticles in aquatic environmants. In Nanoscience and nanotechnology: Environmental and human health implications. Wiley. doi:10.1002/9781444307504.ch6

Holland, J. H. (1992). *Adaptiation in natural and atrificials systems*. MIT Press.

Hollander, A., Scheringer, M., Shatalov, V., Mantseva, E., Sweetman, A., Roemer, M., & van de Meent, D. et al. (2008). Estimating overall persistence and long-range transport potential of persistent organic pollutants. a comparison of seven multimedia mass balance models and atmospheric transport models. *Journal of Environmental Monitoring*, *10*(10), 1139–1147. doi:10.1039/b803760d PMID:18843390

Hutson, N. D., Ryan, S. P., & Touati, A. (2009). Assessment of PCDD/F and PBDD/F emissions from coal-fired power plants during injection of brominated activated carbon for mercury control. *Atmospheric Environment*, *43*(26), 3973–3980. doi:10.1016/j.atmosenv.2009.05.026

Klasmeier, J., Matthies, M., Macleod, M., Fenner, K., Scheringer, M., Stroebe, M., & Wania, F. et al. (2006). Application of multimedia models for screening assessment of long-range transport potential and overall persistence. *Environmental Science & Technology*, *40*(1), 53–60. doi:10.1021/es0512024 PMID:16433332

Kormann, F., Rominger, J., Schowanek, D., Wagner, J.-O., Schroder, R., & Wind, T. et al.. (2006). Modelling the fate of down-the-drain chemicals in rivers: An improved software for GREAT-ER. *Environmental Modelling & Software*, *21*(7), 925–936. doi:10.1016/j.envsoft.2005.04.009

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

Mackay, D. (1979). Finding fugacity feasible. *Environmental Science & Technology*, *13*(10), 1218–1223. doi:10.1021/es60158a003

Mackay, D., Di Guardo, A., Paterson, S., Kicsi, G., Cowan, C. E., & Kane, M. (1996). Assessment of chemical fate in the environment using evaluative, regional and local-scale models: Illustrative application to chlorobenzene and linear alkylbenzene sulfonates. *Environmental Toxicology and Chemistry*, *15*(9), 1638–1648. doi:10.1002/etc.5620150930

Mackay, D., Shiu, W. Y., Ma, K.-C., & Lee, S. C. (2007). *Handbook of physical-chemical properties and environmental fate for organic chemicals*. Boca Raton, FL: Taylor & Francis.

MacLeod, M., Woodfine, D. G., Mackay, D., McKone, T., Bennett, D., & Maddalena, R. (2001). BETR North America: A regionally segmented multimedia contaminant fate model for North America. *Environmental Science and Pollution Research International*, *8*(3), 156–163. doi:10.1007/BF02987379 PMID:11505899

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Martin, D., Maran, U., Sild, S., & Karelson, M. (2007). QSPR modeling solubility of polyaromatic hydrocarbons and fullerene in 1-octanol and n-heptane. *The Journal of Physical Chemistry B*, *111*(33), 9853–9857. doi:10.1021/jp071679x PMID:17661505

McKone, T. (1993). *CalTOX, a multimedia total-exposure model for hazardous-wastes sites part II: The dynamic multimedia transport and transformation model*. Retrieved from http://www.dtsc.ca.gov/AssessingRisk/upload/techman2.pdf

METI. (1973). *Chemical substances control law ministry of economy, trade and industry*. METI.

Mitrano, D. M., Lesher, E. K., Bednar, A., Monserud, J., Higgins, C. P., & Ranville, J. F. (2012). Detecting nanoparticulate silver using single-particle inductively couple plasma-mass spectrometry. *Environmental Toxicology and Chemistry*, *31*(1), 115–121. doi:10.1002/etc.719 PMID:22012920

Mostrag, A., Puzyn, T., & Haranczyk, M. (2010). Modeling the overall persistence and environmental mobility of sulfur-containing polychlorinated organic compounds. *Environmental Science and Pollution Research International*, *17*(2), 470–477. doi:10.1007/s11356-009-0257-7 PMID:19937279

Netzeva, T. I. (2005). Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships - The report and recommendations of ECVAM workshop 52. *ATLA Alternatives to Laboratory Animals, 33*, 155-173.

Nowack, B., Rnaville, J. F., Diomond, S., Gallego-Urrea, J. A., Metcalfe, C., & Rose, J. et al.. (2012). Potential scenarios for nanomaterial release and subsequent alteration in the environment. *Environmental Toxicology and Chemistry*, *31*(1), 50–59. doi:10.1002/etc.726 PMID:22038832

OECD. (2002). *Report of the OECD/UNEP workshop on the use of multimedia models for estimation overall environmental persistence and long-range Transport in the context of PBTs/POPs assessment*. OECD Environment, Health and Safety Publications.

OECD. (2004). *Guidance document on the use of multimedia models for estimatig overall environmental persistence and long-range transport*. Paris, France: Health and Safety Publications.

OECD. (2007). Guidance document on the validation of (quantitative) structure activity relationship [(Q)SAR] models. no. 69. Organisation of Economic Cooperation and Development.

Prinn, R. G., Huang, J., Weiss, R. F., Cunnold, D. M., Fraser, P. J., & Simmonds, P. G. et al.. (2001). Evidence for substantial variations of atmospheric hydroxyl radicals in the past two decades. *Science*, *292*(5523), 1882–1888. doi:10.1126/science.1058673 PMID:11337586

Puzyn, T. (2011). On the replacement of empirical parameters in multimedia mass balance models with QSPR data. *Journal of Hazardous Materials*, *192*(3), 970–977. doi:10.1016/j.jhazmat.2011.05.078 PMID:21741174

Puzyn, T., & Falandysz, J. (2005). Octanol/water partition coefficients of chloronaphthalenes. *Journal of Environmental Science and Health. Part A, Environmental Science and Engineering & Toxic and Hazardous Substance Control*, *40*(9), 1651–1663. doi:10.1081/ESE-200067976 PMID:16134358

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Puzyn, T., & Falandysz, J. (2007). QSPR modeling of partition coefficients and Henry's law constants for 75 chloronaphthalene congeners by means of six chemometric approaches - A comparative study. *Journal of Physical and Chemical Reference Data*, *36*(1), 203–214. doi:10.1063/1.2432888

Puzyn, T., Falandysz, J., Jones, P. D., & Giesy, J. P. (2007). Quantitative structure - activity relationships for the prediction of relative in vitro potencies (REPs) for chloronaphthalenes. *Journal of Environmental Science and Health. Part A, Environmental Science and Engineering & Toxic and Hazardous Substance Control*, *42*(5), 573–590. doi:10.1080/10934520701244326 PMID:17454365

Puzyn, T., Leszczynska, D., & Leszczynski, J. (2009b). Quantitative structure-activity relationships (QSARs) in the European REACH system. Could these apporaches be applied to nanomaterials? In *Practical aspects of computational chemistry. Methods, concepts and applications.* Heidelberg, Germany: Springer. doi:10.1007/978-90-481-2687-3_9

Puzyn, T., Mostrag, A., Falandysz, J., Kholod, Y., & Leszczynski, J. (2009c). Predicting water solubility of congeners. Chloronaphthalenes - A case study. *Journal of Hazardous Materials*, *170*(2-3), 1014–1022. doi:10.1016/j.jhazmat.2009.05.079 PMID:19524360

Puzyn, T., Mostrąg, A., Suzuki, N., & Falandysz, J. (2008c). QSPR-based estimation of the atmospheric persistence for chloronaphthalene congeners. *Atmospheric Environment*, *42*(27), 6627–6636. doi:10.1016/j.atmosenv.2008.04.048

Puzyn, T., Mostrąg-Szlichtyng, A., Suzuki, N., & Haranczyk, M. (2009a). Metody chemometryczne w ocenie ryzyka. Ilościowe zależności pomiędzy strukturą chemiczną a właściwościami (QSPR) dla nowych rodzajów zanieczyszczeń chemicznych. In *Chemometria w nauce i praktyce*. Kraków: Wydawnictwo Instytutu Ekspertyz Sądowych.

Puzyn, T., Rostkowski, P., Swieczkowski, A., Jedrusiak, A., & Falandysz, J. (2006). Prediction of environmental partition coefficients and the Henry's law constants for 135 congeners of chlorodibenzothiophene. *Chemosphere*, *62*(11), 1817–1828. doi:10.1016/j.chemosphere.2005.07.055 PMID:16214201

Puzyn, T., Suzuki, N., & Haranczyk, M. (2008b). How do the partitioning properties of polyhalogenated POPs change when chlorine is replaced by bromine? *Environmental Science & Technology*, *42*(14), 5189–5195. doi:10.1021/es8002348 PMID:18754368

Puzyn, T., Suzuki, N., Haranczyk, M., & Rak, J. (2008). Calculation of quantum-mechanical descriptors for QSPR at the DFT level. Is it necessary? *Journal of Chemical Information and Modeling*, *48*(6), 1174–1180. doi:10.1021/ci800021p PMID:18510372

Puzyn, T., Suzuki, N., & Sakurai, T. (2011). Estimating persistence of brominated and chlorinated organic pollutants in air, water, soil, and sediments with the QSPR-based classification scheme. *Molecular Diversity*, *15*(1), 173–188. doi:10.1007/s11030-010-9250-9 PMID:20386980

Roy, K., Chakraborty, P., Mitra, I., Ojha, P. K., Kar, S., & Das, R. N. (2013). Some case studies on application of "r(m)(2)" 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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Scheringer, M., Wegmann, F., Fenner, K., & Hungerbuhler, K. (2000). Investigation of teh cold condensation of persistanet organic chemicals with global multimedia fate model. *Environmental Science & Technology*, *34*(9), 1842–1850. doi:10.1021/es991085a

Schowanek, D., Fox, K., Holt, M., Schroeder, F. R., Koch, V., & Cassani, G. et al. (2001). GREAT-ER: A new tool for management and risk assessment of chemicals in river basins. Contribution to GREAT-ER. *Water Science and Technology*, *43*, 179–185. PMID:11380178

Schowanek, D., & Webb, S. (2002). Exposure simulation for pharmaceuticals in European surface waters with GREAT-ER. *Toxicology Letters*, *131*(1-2), 39–50. doi:10.1016/S0378-4274(02)00064-4 PMID:11988357

Sun, T. Y., Gottschalk, F., Hungerbuhler, K., & Nowack, B. (2014). Comprehensive probabilistic modelling of environmental emissions of engineered nanomaterials. *Environmental Pollution*, *185*, 69–76. doi:10.1016/j.envpol.2013.10.004 PMID:24220022

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

Toropov, A. A., Leszczynska, D., & Leszczynski, J. (2007). Predicting water solubility and octanolwater partition coefficient for carbon nanatubes based on the chiral vector. *Computational Biology and Chemistry*, *31*(2), 127–128. doi:10.1016/j.compbiolchem.2007.02.002 PMID:17363328

UN-ECE. (1998). *Protocol to the 1979 convention on long range transboundary air pollution on persistent organic pollutants and executive body decision 1998/2 on information to be submitted and the procedure for adding substances to annexes I, II or III to the protocol on persistent organic pollutants. Document. ECE/EB.AIR/60*. New York: United Nations.

UNEP. (2001). *Final act of the conference of plenipotentiaries on the Stockholm convention on persistent organic pollutants. Document. UNEP/POPS/CONF/4*. Stockholm: United Nations Environment Programme.

USEPA. (1999). *Toxic substances control act*. The United States Environmental Protection Agency.

van der Meent, D. (1993). *SimpleBox: A generic multi-media fate evaluation model*. Institute of Public Health and the Environment.

van Wezel, A. P., & Jager, T. (2002). Comparison of two screening level risk assessment approaches for six disinfectants and pharmaceuticals. *Chemosphere*, *47*(10), 1113–1128. doi:10.1016/S0045-6535(02)00048-6 PMID:12137045

van Wijk, D., Chenier, R., Henry, T., Hernando, M. D., & Schulte, C. (2009). Integrated approach to PBT and POP prioritization and risk assessment. *Integrated Environmental Assessment and Management*, *5*(4), 697–711. doi:10.1897/IEAM_2009-034.1 PMID:19552502

Vermeire, T. G., Jager, D. T., Bussian, B., Devillers, J., den Haan, K., Hansen, B., & van der Zandt, P. T. J. et al. (1997). European Union system for the evaluation of substances (EUSES). Principles and structure. *Chemosphere*, *34*(8), 1823–1836. doi:10.1016/S0045-6535(97)00017-9 PMID:9114488

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

von der Krammer, F., Ferguson, P. L., Holden, P. A., Masion, A., Rogers, K. R., & Klaine, S. J. et al.. (2012). Analysis of engineered nanomaterials in complex matrices (environment and biota): General consideractions and conceptual case studies. *Environmental Toxicology and Chemistry*, *31*(1), 32–49. doi:10.1002/etc.723 PMID:22021021

Walczak, B., Daszykowski, M., & Stanimirova, I. (2009). Robust methods in QSAR. In T. Puzyn, J. Leszczynski, & M. T. D. Cronin (Eds.), *Recent advances in QSAR studies. Challenges and advances in computational chemistry and physics*. Dordrecht, The Netherlands: Springer.

Wania, F. (2006). Potential of degradable organic chemicals for absolute and relative enrichment in the *Arctical*. *Environmental Science & Technology*, *40*(2), 569–577. doi:10.1021/es051406k PMID:16468404

Wania, F., & Daly, G. L. (2002). Estimating the contribution of degradation in air and deposition to the deep sea to the global loss of PCBs. *Atmospheric Environment*, *36*(36-37), 5581–5593. doi:10.1016/S1352-2310(02)00693-3

Wania, F., & Mackay, D. (1995). A global distribution model for persistent organic-chemicals. *The Science of the Total Environment*, *160-61*, 211–232. doi:10.1016/0048-9697(95)04358-8

Wegmann, F., Cavin, L., MacLeod, M., Scheringer, M., & Hungerbuhler, K. (2009). The OECD software tool for screening chemicals for persistence and long-range atmospheric transport potential. *Environmental Modelling & Software*, *24*(2), 228–237. doi:10.1016/j.envsoft.2008.06.014

Zukowska, B., Breivik, K., & Wania, F. (2006). Evaluating the environmental fate of pharmaceuticals using a level III model based on poly-parameter linear free energy relationships. *Environmental Science & Technology*, *359*, 177–187. PMID:16009398

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

KEY TERMS AND DEFINITIONS

Applicability Domain: Describes a theoretical space determined by the values of **X** (matrix of descriptors) and **y** (end-point), within which the predictions made by QSPR/QSAR models are reliable.

ConGENER: Software that generate a series of internal co-ordinate matrices for all probable congeners on the basis of a mother (other than substituted) molecule internal coordinate matrix which indicates hydrogen atoms likely to be substituted.

EPI Suite: A tool widely applied for the purpose of risk assessment in the case of new chemical substance to estimate its environmental fate.

Molecular Descriptors: Mathematic transformation (encode) of chemical structures into numbers.

Multimedia Mass-Balance Models: Research tools ensuring the studies into the impact of individual environmental components on the transport and deposits of chemical substances.

OECD: (The Organisation for Economic Co-operation and Development) International organisation established to simulate economic progress and world trade, associating of 34 countries. The organisation was founded in 1961.

Persistence Organic Pollutants: Chemicals, such as chloro- and bromo-substituted biphenyls, naphthalenes, dibenzo-*p*-dioxins, dibenzofurans, and diphenyl ethers that have been identified in almost all environmental compartments. Due to their high liphophilicity and resistance to naturally occurring degradation processes, they are prone bioaccumulation in human and animal tissues. In the organism, they are capable to induce various toxic effects, including carcinogenicity, reproductive disorders related to disrupting the hormonal system, immunotoxicity and damages to the central and peripheral nervous systems. They are also suspected to be responsible for the increasing number of patients nowadays suffering from allergies and hypersensitivity.

REACH Regulations: Regulation, Evaluation, Authorisation and Restriction of Chemicals is the European Union regulations to estimate and/or decrease the potential impacts of chemicals on both human health and the environment.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 13
QSAR-Based Studies of Nanomaterials in the Environment

Valeria V. Kleandrova
University of Porto, Portugal

Alejandro Speck-Planche
University of Porto, Portugal

Feng Luan
Yantai University, China & University of Porto, Portugal

M. Natália D. S. Cordeiro
University of Porto, Portugal

ABSTRACT

Nanotechnology is a newly emerging field, posing substantial impacts on society, economy, and the environment. In recent years, the development of nanotechnology has led to the design and large-scale production of many new materials and devices with a vast range of applications. However, along with the benefits, the use of nanomaterials raises many questions and generates concerns due to the possible health-risks and environmental impacts. This chapter provides an overview of the Quantitative Structure-Activity Relationships (QSAR) studies performed so far towards predicting nanoparticles' environmental toxicity. Recent progresses on the application of these modeling studies are additionally pointed out. Special emphasis is given to the setup of a QSAR perturbation-based model for the assessment of ecotoxic effects of nanoparticles in diverse conditions. Finally, ongoing challenges that may lead to new and exciting directions for QSAR modeling are discussed.

INTRODUCTION

The 21th century has witnessed a fast-dynamic growth of the nanotechnology field, leading in turn to the development of an innumerable number of nanomaterials with unique properties and a variety of applications in many sectors. However, the large-scale production and increasingly likelihood exposure to novel nanomaterials has spurred great concern regarding their potential adverse effects to humans and the environment. This chapter will provide an overview of *in silico* approaches developed so far for the prediction of nanomaterials environmental toxicity; from the use of classical approaches forward-

DOI: 10.4018/978-1-4666-8136-1.ch013

Copyright © 2015, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

establishing Quantitative Structure−Activity Relationships (QSAR) to the applications of more evolved computational models. Particular attention will be paid to the setup of a new *in silico* perturbation-based model aimed at reliable predicting the ecotoxicity of metal-based nanoparticles under a wide range of experimental conditions.

NANOMATERIALS AND ENVIRONMENTAL RISKS

The idea that launched the new field nanotechnology started with the famous lecture "There's plenty of room at the bottom" given by physicist Richard Feynman at the Annual Meeting of the American Physical Society in 1959 (Feynman, 1960). In that meeting, Feynman wondered: "What would the properties of materials be if we could really arrange the atoms the way we want them?", foreseeing the production of materials at the nanometer scale with promising technical, industrial and biological applications. But the term "nanotechnology" was only coined decades later, firstly by Norio Taniguchi in 1974 within his inspections of ultra-precision machining (Taniguchi, 1974), and then by K. Eric Drexler in his 1986 book *Engines of Creation: The Coming Era of Nanotechnology* (Drexler, 1992). Though the real burst of nanotechnology didn't come until the early 1990s, with the development of sophisticated techniques for the characterization and manipulation of individual atoms such as scanning probe microscopy.

Nanomaterials are usually considered to entail materials that have single units with at least one dimension between 1 nm and 100 nm (Oberdörster et al., 2007). Nanomaterials can additionally be classified as one-dimensional (1-D), two-dimensional (2-D) and three-dimensional (3-D), according to the number of dimensions that are confined to the nanoscale range (< 100 nm). 1-D materials include nanofilms, nanolayers, and nanocoatings, while 2-D include nanotubes, nanorods, and nanowires. Finally, the most common 3-D materials are nanoparticles (NPs), the building blocks for nanotechnology applications. Nanomaterials can be amorphous or crystalline, be composed of single or multi-chemical elements, and exhibit various shapes and forms.

The advances in nanotechnology have made possible the development of many new materials and devices with a vast range of applications in medicine, cosmetics, electronics, biomaterials and energy production. Indeed due to the rapid pace at which nanotechnology is evolving and expanding into society via its applications, it is expected that many new innovative products containing manufactured nanomaterials will continue to vastly grow. Commonly, manufactured nanomaterials fall generally into two categories: carbon-based and metal-based materials (Bréchignac et al., 2007). Regarding carbon-based nanomaterials, these usually involve diamond, carbon-nanotubes, graphene and fullerenes, the latter being a class of allotropes of carbon that conceptually are graphene sheets rolled into tubes or spheres. Most of these nanomaterials are of interest for their mechanical strength and electrical properties (De Volder et al., 2013). Moreover, fullerenes in general are under study for potential medical applications, ranging from specific binding of antibiotics on resistant bacteria to even target certain types of cancer cells such as melanoma (Bakry et al., 2007; Simon et al., 2007). Carbon nanotubes for instance, have already been used as drug, gene, or protein carriers (Bianco et al., 2005). Nanoscale materials are being further developed in therapeutic modalities for advanced selective drug delivery systems and targeted nanotherapy (Ravi Kumar, 2000). Many of these modalities include polymeric NPs (Stupp et al., 2005), micelles (Kim et al., 2005), liposomes and dendrimers (Popat et al., 2007; Portney & Ozkan, 2006), hydrogels (Vinogradov et al., 2002), nanoshells (Zahr et al., 2005), and smart surfaces. The key performance characteristics of these materials are the high loading capacity,

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

release kinetics, circulation time, bio-distribution, size distribution, and stability (Allen et al., 1999). Furthermore, a recently published review shows that different carbon-based nanomaterials, including nanodiamonds, fullerenes, carbon nanotubes, graphene, carbon nanofibers, carbon nanocones/disks and nanohorns, as well as their functionalized forms, can be used in sample preparation processes in analytical chemistry (Zhang et al., 2013). With regard to metal-based nanomaterials, these include metal and metal oxide NPs based on titanium, zinc, silver, copper, aluminium or cerium among others. For instance, oxides TiO_2 and ZnO have been used in cosmetics, sunscreens, solar-driven self-cleaning coatings, and textiles for enhanced product transparency, as well as in paintings, water treatment and food (Cai et al., 2006; Serpone et al., 2007; Yuranova et al., 2007). As to nanosized silver, it is used today in an increasing number of consumer and medical products. Typical examples include band aids, cosmetics, food complements, anti-odour textiles, dental materials, bone implants, water disinfectants, as well as antibacterials, antivirals, etc. agents (Rai et al., 2014; Wijnhoven et al., 2009). On the other hand, CuO nanocrystals have been found that could replace noble metals in newer catalytic converters for the car industry (Zhou et al., 2006). Further, metal salts including CdSe quantum dots and iron (Fe) NPs have been exploited in many practical applications related to electronics and biomedical imaging techniques (Alivisatos, 2004; Chen & Gerion, 2004). Other specialized industrial applications of nanomaterials involve for instance aluminium (Al) NPs as fuel additives and MnO as a catalyst or in battery technology (Schrand et al., 2012).

Manufactured nanoparticles (MNPs) are receiving a great deal of attention lately both from the basic sciences and industry, as well as from the public at large, since existing and potential applications are almost endless. But along with the benefits, the use of MNPs raises many questions and concerns due to the lack of knowledge regarding their health and environmental impacts (Borm et al., 2006; Olson & Gurian, 2012; Schrand et al., 2012; Walker & Bucher, 2009; Warheit et al., 2008). In fact, uncertainties regarding those are high since the properties exhibited by NPs are often completely different from those of bulk forms, affecting their physicochemical and biological behavior, which may well result in greater toxicity. At the same time, the use, production and disposal of nanoscale materials raises concerns about nano-pollution, taking into account that NPs can be released in the air or water during production or as waste by-product of production, and ultimately accumulate in the soil, water or plant life, endangering the health of living organisms and ecosystems. There is thus an increasingly need to address the overall environmental impact of released NPs to support the development of a sustainable and safe nanotechnology.

IN SILICO STUDIES OF NANOMATERIALS IN THE ENVIRONMENT

As can be judged, while prompt development of nanomaterials will certainly continue, the application of these materials might be limited by their potential health-risks and environmental impacts (Bar-Ilan et al., 2009; García et al., 2011; Gong et al., 2011; Griffitt et al., 2008; Heinlaan et al., 2008; Hund-Rinke & Simon, 2006; Kasemets et al., 2009; Lin & Xing, 2007; Ma et al., 2011; Ma et al., 2010; Marsalek et al., 2012; Sadiq et al., 2011; Zhu et al., 2010; Zhu et al., 2012; Zhu et al., 2009; Zhu et al., 2008). And indeed, there have been already several reports on the deleterious effects of nanomaterials to human health and ecosystems (Arora et al., 2012; Bussy, C. et al., 2013; Liu et al., 2013; Schrand et al., 2012; Tsoy et al., 2013). For example, carbon nanotubes have been shown to produce oxidative stress and being just as toxic as asbestos (Liu et al., 2013; Schrand et al., 2012).

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Similarly to any other manufactured chemicals, toxicity evaluation studies of nanomaterials have been carried out on a few cell lines or species that have been accepted by regulatory agencies as models for defining toxicological effects. As a rule, an initial screen of toxicity is accomplished by resorting to *in vitro* methodologies, even though the bridge between *in vitro* and *in vivo* remains. Nevertheless, *in vivo* mimicry of higher order organisms is difficult due to the complex nature of biological interactions, and often insurmountable differences between laboratory animals and humans do exist (Eisenbrand et al., 2002; Oberdorster, 2010; Schrand et al., 2012; Shvedova & Kagan, 2010). Alternatively, since *in vivo* experiments are expensive, slow and ethically questionable, low-cost *in vitro* assays without reducing the efficiency and reliability of the risk assessment are usually carried out. One should notice here that some test models and standard testing protocols applied for non-nanoform substances may not necessarily be appropriate or optimal for the testing of MNPs, but nevertheless ongoing efforts in the research community are currently addressing such issues (Holden et al., 2013; Liu et al., 2013; Nel et al., 2013; Schrand et al., 2012).

To date, many important aspects regarding the ecotoxicity profiles of NPs are not well understood, and sometimes experimental assays carried out by different groups on the same NPs have showed a remarkable lack of reproducibility and consistency (Lanone et al., 2009). Powerful experimental techniques focused on high-throughput/content screening (HTS/HCS) could be essential for the fast and efficient evaluation of many toxic effects of NPs (Holden et al., 2013). Nevertheless, even with the use of HTS/HCS, there is a very low probability of covering and filtering the huge "nanotoxicity space". This comes from the fact that there are many combinations of factors such as chemical compositions, sizes and conditions under which the sizes were determined, shapes, coating agents, measures of the toxic effects, living systems (cell lines, crustaceans, algae, bacteria, fungi, plants, fish, etc.), and times at which living systems were exposed to NPs. Computational methods may offer a magnification of the 'sweet spot' regarding the virtual screening of NPs for prediction of toxicity profiles, helping to rationalize the experimental assays by diminishing time and financial resources. Among these methods, Quantitative Structure–Activity Relationships (QSAR) models have been used as promising tools for modeling different biological effects of chemicals (Estrada et al., 2004; Estrada et al., 2006; Estrada et al., 2000; Gonzalez-Diaz et al., 2014; Luan et al., 2013; Prado-Prado et al., 2011; Speck-Planche & Cordeiro, 2014; Speck-Planche et al., 2013a; Speck-Planche et al., 2013b; Tenorio-Borroto et al., 2014), including diverse toxicity profiles (Cruz-Monteagudo et al., 2013; Cruz-Monteagudo et al., 2008; Cruz-Monteagudo et al., 2006; Estrada et al., 2001; Estrada & Uriarte, 2001; Estrada et al., 2003; Helguera et al., 2007; Morales Helguera et al., 2008; Speck-Planche et al., 2012; Tenorio-Borroto et al., 2013). In the field of nanotechnology, QSAR models have provided deeper insights regarding the most relevant aspects of NPs (Poater et al., 2009). Nevertheless, only a few studies have been carried out thus far aiming at the *in silico* prediction of nanotoxicity by resorting to QSAR models (Fourches et al., 2010; Kar et al., 2014a; Kar et al., 2014b; Liu et al., 2011; Liu et al., 2013; Puzyn et al., 2011; Toropov et al., 2012). From now on, this chapter will be devoted to summarize the main progresses in the computational risk assessment of nanomaterials. Towards that goal, prediction of both cytotoxicity (against mammalian cell lines) and ecotoxicity (involving bio-indicator species such as bacteria, algae, fungi, crustaceans, fish, and plants) will be discussed because of their potential direct or indirect implications in environmental risk assessment.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Classical QSAR Models for Prediction of Toxic Effects of Nanoparticles

In general terms, QSAR approaches aim to establish mathematical models disclosing possible parallelisms/relationships between a set of predictor variables *X*, known as molecular descriptors, and the response variable *Y*, which describes a certain type of biological effect (Hansch & Leo, 1995). From one side, molecular descriptors can be viewed as the final result of a defined mathematical procedure that transforms chemical information encoded within a symbolic representation of a molecule into a useful number or the result of some standardized experiment (Todeschini & Consonni, 2000). On the other hand, the biological effect can be associated with the inhibitory activity of a chemical against different types of targets (biomacromolecules, microorganisms, cell lines, etc), the toxicological profile, or any property related with the absorption, distribution, metabolism, and elimination (ADME). QSAR models are obtained by means of statistical methods (van de Waterbeemd, 1995), which include regression techniques such as (but not limited to) multiple linear regression (MLR) and partial least squares (PLS), or classification techniques such as (but not limited to) linear discriminant analysis (LDA). In addition, powerful machine learning methods like artificial neural networks (ANN), support vector machines (SVM), random forest (RF), consensus clustering (CC) and many others can be employed (Hill & Lewicki, 2006).

Classical QSAR models are based on assumptions, which describe the structural dependence of biological activities (or any other biological effect) either by local physicochemical descriptors such as hydrophobicity, molar refractivity, and electronic factors (Hansch analysis), or by indicator variables (discrete descriptors) encoding different structural features (Free-Wilson analysis) (Kubinyi, 1993). Current classical QSAR models maintain the same principles in the way that small datasets of structurally related chemicals are used. However, other groups of local and global molecular descriptors such as those derived from quantum-mechanics calculations or exhibiting graph-theoretical nature have been widely employed (Todeschini & Consonni, 2009).

Several works focusing on predicting nanotoxicity have been reported in the scientific literature, in which the fundamental principles of QSAR studies have been adapted to the computational assessment of different biological effects of NPs. The first of these works was inspired by the idea that the evaluation of biological effects of MNPs is extremely important for nanotechnology sustainability, and there is thus an urgent need for efficient *in silico* approaches able of predicting biological effects of MNPs. Therefore, the authors developed statistical significant QSAR models by establishing relationships between the toxicities of nanomaterials and their corresponding physical, chemical, and geometrical properties, either measured experimentally or computed from the structure of MNPs (Fourches et al., 2010). Towards that end, two representative sets of MNPs involving *in vitro* cell-based assays have been employed. The first set was formed by 51 various MNPs with diverse metal cores (Shaw et al., 2008), and 44 of them were used to derive the nano-QSAR models. Each NP was tested against 4 cell lines, employing 4 doses, and through 4 different assays (64 responses of biological activity). The authors condensed the entire biological effects (cytotoxicities) of each NP into one average response. At the same time, experimental properties such as size, relaxivity, and zeta potential were used as descriptors for the MNPs. Afterwards the SVM technique was used to develop the QSAR models. The whole dataset was split 5 different times, creating 5 different training (80%) and prediction (20%) sets. The best model yielded an external accuracy of 73%. In the case of the second dataset, 109 MNPs with the same core (cross-linked iron oxide with

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

amine groups) but decorated with different organic molecules (Weissleder et al., 2005) was studied. The uptake of the NPs in PaCa2 cells was used as biological response, while 150 different computational molecular descriptors were used to characterize the chemical structures of the different molecules in the surface of the MNPs (coating agents). Then, QSAR models based on the *k*-nearest neighbors (*k*NN) regression technique were setup, and the best model displayed a reasonable external prediction power ($R^2 = 0.72$). The results of this work clearly demonstrate that nano-QSAR models can be used for two main purposes: prediction of the biological activity profiles of novel nanomaterials, and prioritization in the design and manufacturing of nanomaterials towards better and safer products.

As a continuation of the work involving the last dataset mentioned above, *i.e.*, that formed by 109 magnetofluorescent NPs, Roy and coworkers built up a new nano-QSAR model using the MLR technique (Kar et al., 2014a). For deriving the model, a pool of 307 descriptors was calculated using different computer programs. The best model contained 6 molecular descriptors, and was rigorously validated internally and externally according to the guidelines established by the Organization for Economic Cooperation and Development (OECD). Accordingly, the values of R^2 and Q^2_{LOO} for the training set (89 cases) were 0.806 and 0.758 respectively, while for the external set (20 cases) R^2_{pred} had a value of 0.879 with a $R^2_{m(test)Scaled} = 0.793$. In this work, the applicability domain was also determined in an attempt to confirm the reliability of the predictions (Dimitrov et al., 2005; Dragos et al., 2009; Fechner et al., 2010; Kulkarni & Zhu, 2008; Netzeva et al., 2005; Weaver & Gleeson, 2008). Additionally, molecular fragments present in the different coating agents were analyzed, and their contributions were calculated. The main purpose underpinning the calculation of fragments contributions was to study the most important molecular features associated with the highest (or the lowest) values of cellular uptake of NPs. One should emphasize here that even when the calculation of quantitative contributions of fragments to different biological effects have been widely reported by different authors (Estrada et al., 1998; Estrada & Peña, 2000; Estrada et al., 2000; Helguera et al., 2005; Helguera et al., 2007; Prado-Prado et al., 2010; Speck-Planche & Cordeiro, 2013), this work introduces for the first time essential and rich information *via* the attained fragment contributions, expanding the horizons for risk assessment in the context of virtual screening of relevant NP libraries.

Another nano-QSAR model has also been developed by employing periodic table-based descriptors to encode cytotoxicity profile of metal oxide NPs (Kar et al., 2014b). In such work, metal electronegativity (χ), the charge of the metal cation corresponding to a given oxide (χ_{ox}), the atomic number, and the valence electron number of the metal were used as simple descriptors to setup QSAR models aiming at predicting the cytotoxicity of metal oxide NPs against the bacterium *Escherichia coli*. As the measure of cytotoxicity, the effective concentration at 50% (EC_{50}) was employed. Then, MLR and PLS were used as statistical regression methods to search for the best QSAR models, which could account for at least 82% of the variance of the cytotoxicity in the training set, with $Q^2_{LOO} \geq 0.75$. In the prediction set, an average R^2_{pred} for the ten derived models (due to random splits) was 0.78 (or higher). Many other statistical indices were calculated, corroborating the quality and predictive power of the nano-QSAR models. A prominent aspect of the latter research work is that the descriptors used to characterize the metal oxide NPs can be obtained easily and quickly from the molecular formula (composition) and the information retrieved from the periodic table. On the other hand, the authors demonstrated that a simple but useful descriptor such as χ_{ox} can efficiently encode the cytotoxicity of metal oxide NPs, leading to models with clear interpretability. Lastly, the authors also hypothesized the possible mechanisms of action of the NPs. Finally, this work served to uncover the fact that the bacterium *Escherichia coli* can be used as a useful target for the study, analysis and prediction of cytotoxicity of NPs. Indeed, two more works have been

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

published previously, and they were focused on modeling the same phenomenon: cytotoxicity of metal oxide NPs against *Escherichia coli* (Puzyn et al., 2011; Toropov et al., 2012). The first of these works was based on the calculation of descriptors derived from quantum-mechanics calculations (Puzyn et al., 2011), while the second one used SMILES-based optimal descriptors, which were calculated from the CORAL software (Toropov et al., 2012). In both cases, the MLR technique was used to derive QSAR models with excellent performance when compared with the research work referred to above.

Another interesting report focused on the computational assessment of cytotoxicity of NPs by means of nano-QSAR models based on statistical classification methods (Liu et al., 2011). Towards that goal, the authors built up a model to predict the cytotoxicity of 9 metal oxide NPs against bronchial epithelial cells (BEAS-2B), which were exposed to different concentrations of the NPs (0.375-200 mg/L) during a period of time of 24 hrs. The study was realized by using cytotoxicity data from a high-throughput screening assay that was processed to identify and label toxic (through the propidium iodide uptake of BEAS-2B cells) versus endpoints relative to an unexposed control cell population. A total of 14 different physicochemical descriptors were used as input variables. The best nano-QSAR classification model based on logistic regression contained the following descriptors: atomization energy of the metal oxide, period of the metal in the NP, and NP primary size, in addition to NP volume fraction (in solution). This model had 100% classification accuracy in both internal and external validations. Thus, one can undoubtedly infer that nano-QSAR models based on classification tasks can be very useful for risk assessment of NPs. Furthermore, when classification models are used, there is no need to predict the exact value of biological effect of NPs. Consequently, the uncertainty of data can be better controlled.

Finally, a research work dealing with another nano-QSAR classification model has been reported in the last year (Liu et al., 2013). The generation of such a model was inspired by the idea regarding the integration of multi-parametric toxicity/bioactivity assays to arrive at statistically meaningful class definitions, that is, bioactivity/inactivity endpoints. Another purpose of that work was to determine and analyze the decision boundaries and the applicability domain. Towards that end, by applying diverse statistical and machine learning methods, different nano-QSAR models were developed from a set of 44 iron oxide core NPs used in molecular imaging and nano-sensing. These NPs contained bioactivity profiles for four cell types and four different assays. Class assignments were pursued by means of bioactivity identification analysis and self-organizing map based consensus clustering, which permitted the setup of nano-QSAR classification models with acceptable performance (classification accuracy >78%). This is, a naive Bayesian classifier (NBC) containing spin-lattice relaxivity and zeta potential as descriptors was employed. The authors successfully established an applicability domain by using the concept of probability density, and decision boundaries were determined for the above NBC by considering different acceptance levels of false negative to false positive predictions.

Despite the fact that the nano-QSAR models discussed above convey great advances for the regulatory risk assessment of diverse biological effects of NPs, several major drawbacks can be noticed. From one side, some of these models use only experimental descriptors, and taking into account that measurements overall embody experimental errors, this may introduce a great uncertainty when the QSAR models are built up (Fourches et al., 2010). On the other hand, whole QSAR models up to now have been based on limited datasets of NPs, in which toxicity tests have been accomplished by considering only one living system (bio-indicator species, cell line, etc.) and only one reported measure of toxic effects. Additionally, such QSAR models did not provide information regarding important factors such as the shape and medium under which the sizes of the NPs have been measured, as well as other experimental conditions such as the time during which the living systems have been exposed to the NPs. Finally, we would like

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

to emphasize that the previously developed QSAR models in nanotechnology have generally focused on the prediction of cytotoxicity, and thus, no model has been reported so far with the aim of performing virtual screenings of ecotoxic effects of NPs. Therefore, the setup of a QSAR model able to predict the ecotoxicity of NPs by considering different experimental conditions would be of great importance in nanotoxicology, an in general terms, for nanotechnology. Bearing in mind all these ideas, and the fact that many NPs are released into the environment in their bare forms due to processes such as manufacture, manipulation and use, we dedicate the next section of this chapter to the analysis of QSAR-perturbation models which are focused on the simultaneous prediction of dissimilar ecotoxic effects of uncoated NPs under multiple experimental conditions.

Perturbation Models: A New Paradigm in QSAR Studies

Recently, it has been reported the possibility to apply perturbation theory to QSAR models (Gonzalez-Diaz et al., 2013). In this context, several QSAR-perturbation models have been developed and thoroughly validated, forward-allowing the prediction of different biological effects of chemicals species in diverse fields such as organic synthesis, computational pharmacokinetic modeling, and the study of process of self-aggregation of drugs into micellar NPs. Taking into consideration the importance of QSAR-perturbation models, we will summarize their main aspects.

These innovative models allow the integration of different kinds of chemical and biological data. One of the principal aspects of QSAR-perturbation models is the use of ontologies, which can be viewed as experimental conditions, *i.e.*, the results of the combinations of different elements or factors. For instance, the ontology c_j can be expressed in the form c_j => (b_t, p_s, d_m, t_e), where b_t characterizes the different biological targets (assay organism, cell line, etc.), p_s describes the diverse shapes (morphologies) of the NPs, d_m represents the conditions under which the sizes of the NPs have been measured (in aqueous solution, culture medium, or for the dry powder), and t_e refers to the different periods of time during which the biological targets have been exposed to the NPs.

The main purpose of the QSAR-perturbation models is to describe two principal aspects. From one side, the model should be sensitive to the variation in both the chemical composition of the NPs and the different sets of experimental conditions. But if the descriptors originally calculated (or experimentally determined) for NPs are used, they will not be able to differentiate the toxic effect of the NPs whenever the experimental condition c_j is modified. In an attempt to accomplish this task, the moving average approach (MAA) can be applied (Hill & Lewicki, 2006), and new descriptors can be generated as follows:

$$DX_i(c_j) = X_i - X_i(c_j)_{avg} \quad (1)$$

In the previous equation, X_i is a defined descriptor or physicochemical property, and $X_i(c_j)_{avg}$ is calculated as the average of all the X_i values, which were tested under the same experimental condition c_j. Equation 1 should be applied to each element of the ontology c_j. In any case, only descriptors of type $DX_i(c_j)$ are employed because they take into account both the chemical composition (or any other property) and the experimental condition c_j under which cases/NPs have been tested.

The second aspect focuses on overcoming the limitations of current classical QSAR models. The main hypothesis assumed by such models is that modifications in the structure and/or composition of the chemicals will produce changes in the activities/toxicities with respect to a hypothetical reference

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

chemical (intercept). This assumption is simple but very weak, due to the fact that, in principle, even more than one chemical could be used as reference at the same time. To overcome such assumption, let us first consider a set of NPs along with their respective toxicities [$TE_i(c_j)$] and descriptors $DX_i(c_j)$. If we generate pairs of NPs, then we can consider the following equation:

$$TE_i(c_j)_{nw} = a_0 \cdot TE_i(c_j)_{rf} + \sum b_j \cdot \left[DX_i(c_j)_{nw} - DX_i(c_j)_{rf}\right] + d_0 \quad (2)$$

where for each pairs of NPs, $TE_i(c_j)_{nw}$ is the toxic effect a defined NP in the new state (experimental condition to be predicted), while $TE_i(c_j)_{rf}$ is the toxic effect of the NP of the same pair selected as reference. The term in Equation 2 involving the descriptors of type $DX_i(c_j)$ can be viewed as the perturbation term [$DDX_i(c_j)$], which expresses the differences between the two NPs in terms of chemical compositions (physicochemical properties), sizes, shapes, biological targets, conditions under which the sizes have been measured, and times during which the biological targets have been exposed to the NPs. Also, the a_0, b_j, and d_0 coefficients can be determined by resorting to regression or classification statistical techniques. Notice that regression will be used if $TE_i(c_j)_{nw}$ and $TE_i(c_j)_{rf}$ are the exact values of toxic effects of the NPs. If these two variables just mentioned are instead categorical variables (defining classes according to certain cutoff values), then, classification methods are to be used.

The application of perturbation theory to nano-QSAR problems is very recent. For this reason, only two works have been reported up to date. The first of these works was devoted to the prediction of the cytotoxicity of NPs under different experimental conditions (Luan et al., 2014). Therein, the authors developed a QSAR-perturbation model aiming at simultaneously predicting the cytotoxicity of 41 different NPs by considering assays against at least 1 out of 15 mammalian cell lines and within 1 out of 5 exposure times. From the physicochemical point of view, the NPs exhibited at least 1 out of 5 possible shape labels, and their sizes were measured in at least 1 out of 4 conditions. The measure of biological effect was TC_{50}, which is the concentration of NPs leading to 50% reduction in cell viability. The NPs were assigned to 1 out of 2 possible classes named "non-cytotoxic" and "cytotoxic". Thus, the NPs were only considered as non-cytotoxic for $TC_{50} \geq 503.572$ μM, otherwise, they were considered as cytotoxic. The descriptors used were the following: molar volume, electronegativity, polarizability, and NP size. These were then transformed to descriptors of the type $DX_i(c_j)$ (Equation 1). Afterwards, by applying the perturbation approach (Equation 2), the authors generated a database of 1681 cases (pairs of NPs), which was randomly split into training (1261 cases) and prediction (420) sets. Then after, LDA was used as the statistical technique to setup the QSAR-perturbation model. The accuracy of the model was higher than 93% for both training and prediction sets. The practical applicability of the model was demonstrated by predicting the cytotoxicity of different NPs such as those formed by silica (SiO_2), nickel (Ni), and nickel(II) oxide (NiO). Each NP was predicted 41 times, and the results of the predictions converged with the reports of the literature. This work represents a great advance in the computational risk assessment of NPs, because, to the best of our knowledge, the present research is the first attempt to simultaneously predict the cytotoxicity of NPs under multiple experimental conditions by applying a single unified QSAR model.

The second work was based on a very similar QSAR-perturbation methodology, but with a different purpose. In this sense, the objective therein was to predict the ecotoxicity of NPs under dissimilar sets of experimental conditions (Kleandrova et al., 2014). Thus, 85 initial NPs were retrieved from the

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

literature, and these have been tested by at least 1 out of 4 measures of ecotoxicity, against at least 1 out of 20 assay organisms (bio-indicators), and within 1 out of 9 intervals of time of exposure. The different NPs displayed at least 1 out of 7 labels related to their shapes, and their sizes were measured in at least 1 out of 5 conditions. The same physicochemical properties employed on the mentioned previous work were used. The QSAR-perturbation model was derived from a database containing 5520 cases, and its accuracy was *ca*. 99% in both training and prediction sets. Three different nickel-based (Ni) NPs with experimental values reported in the literature were predicted 85 times. As in the case of the first work referred to above, the predictions were also found to be in very good agreement with the experimental evidences, confirming that Ni NPs are not ecotoxic when compared with other NPs. The results of this study thus provided a valuable tool towards an efficient multi-tasking prediction of one of the most important aspects of risk assessment of NPs: *ecotoxicity*. Until now, this work constitutes the only report that really deals with the ecotoxicological profiles of nanomaterials.

The two QSAR-perturbation models reported here have a common detail that is of outstanding importance, since another major limitation of classical QSAR models is overcome. This is related with the fact that in classical models, each endpoint-case to be classified – if the dataset comprises only a single prediction to be made, as well as any future decision to classify a chemical species (external prediction) are based on the reliability of the predictions according to certain criteria belonging to the applicability domain (Dimitrov et al., 2005; Dragos et al., 2009; Fechner et al., 2010; Kulkarni & Zhu, 2008; Netzeva et al., 2005; Weaver & Gleeson, 2008). In the specific case of the QSAR-perturbation models, random pairs of NPs were generated, in which each NP was used many times in both conditions, that is, as one NP in the initial state (reference NP) and as one NP in the new (output or final) state. Therefore, each NP used as reference participated in many predictions of other NPs, and simultaneously, the very same NP was predicted many times using the others. This procedure breaks the boundaries of the classical QSAR models, allowing one to obtain more precise and realistic results. On the other hand, the decision to classify NPs using the QSAR-perturbation models is based on consensus predictions. For this reason, the new (external) NPs were predicted 41 times in the first work referred, and 85 times in the second one, since those are the number of NPs which were originally used to setup the QSAR-perturbation models. Thus, any NP is correctly predicted/classified according to its experimental value of toxicity (compared with the cutoff values) when the consensus percentage is higher than 50%. For a percentage lower than 50% the prediction is considered to fail, while a percentage equal to 50% is considered to be a random prediction. Notice that the consensus percentage is the ratio of the number of times that a NP was correctly classified with respect the total number of possible predictions (*i.e.*, the number of NPs originally used in the dataset).

A QSAR PERTURBATION-BASED MODEL FORWARD-TARGETING ECOTOXITICY OF METAL-BASED NANOPARTICLES

Here, one should emphasize that often no strong linear relationships between the descriptors of the chemical species and their biological activities/toxicities are found. One of the principal motives is that sometimes the process or phenomenon understudy is very complex. In such cases, machine learning methods like ANN can be used because they have become useful tools for virtual screening (Prado-Prado et al., 2010; Speck-Planche & Kleandrova, 2012; Tenorio-Borroto et al., 2012). Indeed, even when

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

simple statistical methods yield good results, ANN can be employed in order to coverage more chemical space in terms of predictive capability. As a contribution for the advancement of our understanding on ecotoxicological profiles of NPs, we present here for the first time, a QSAR-perturbation model based on ANN aimed at predicting the ecotoxicity of metal-based NPs by taking into account the following varying factors: chemical composition, size, conditions under which the size is measured, shape, type of measures of the ecotoxicity, biological targets (bio-indicators), and time during which the biological targets are exposed to the NPs.

Dataset, Descriptors, and Set-Up of the QSAR-Perturbation Model

The dataset employed for setting the current QSAR model was the same as that discussed in the second report focused on establishing a QSAR-perturbation model (Kleandrova et al., 2014). So we will limit ourselves here to a brief outline highlighting only the most important aspects. Each of these 85 NPs were assayed by considering at least 1 out of 4 measures of ecotoxicity (m_e), against at least 1 out of 20 biological targets/bio-indicator species (b_t), where NPs exhibited at least 1 out of 7 types of shape labels (p_s). At the same time, the NP size was measured in at least 1 out of 5 conditions (d_m), and at least 1 out of 9 different times (t_e) during which biological targets were exposed to NPs in the ecotoxicity tests.

In the dataset, all the 85 cases/ NPs were associated to two possible categories related with its ecotoxicity in a specific experimental condition c_j. Each of the NPs was assigned to the category "non-ecotoxic" [$ETOX_i(c_j) = +1$] when it had a large value of the measure of ecotoxicity, otherwise, the NP was assigned to the category "ecotoxic" [$ETOX_i(c_j) = -1$]. We need to point out here that $ETOX_i(c_j)$ is a binary (categorical) variable which is employed to classify non-ecotoxic or ecotoxic NPs, according to certain cutoff values of ecotoxicity (Table 1). Another important element is the use of the experimental condition c_j, embodying the ontology with the form c_j => (m_e, b_t, p_s, d_m, t_e). That is, the combination of these five elements describes a specific experimental condition. Each NP was characterized by four descriptors. The first three were the molar volume (V), the electronegativity (E), and the polarizability (P), thus based on the chemical composition of the different NPs. These physicochemical descriptors were retrieved from the public source Chemicool Periodic Table (Hsu, 2013). In the case of the NPs composed by more than one element, the descriptors mentioned above were normalized; that is, the descriptors for such kind of NPs were computed as the sum of the descriptors of all the atoms forming the NPs, and the result of the sum was divided by total number of atoms. The fourth descriptor was the size L (expressed in nm) of each NP, which has been experimentally measured. All chemical and ecotoxicological data were deposited in an excel file, shown in Supporting Information (file **SM1**).

As explained before, Equation 1 was applied to calculate the descriptors of the type $DX_i\left(c_j\right)$, and $X_i\left(c_j\right)_{avg}$ for the entire set n_j was calculated as the average of all the X_i values for only those cases/NPs that were considered as non-ecotoxic cases (subset of n_j) in the same element of c_j. Then after, the 85 NPs were used to setup 5939 random pairs of NPs. Notice now that Equation 2 expresses the perturbation approach when there is a linear dependence between the toxic effect of the NP in the new state and the toxic effect of the NP in the reference state together with the perturbation terms. A more general expression for any type of dependence would be:

$$ETOX_i(c_j)_{nw} = f[ETOX_i(c_j)_{rf}, DDX_i(c_j)] \tag{3}$$

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 1. Cutoff values used for the different measures of ecotoxicity

Measure of Ecotoxic Effect	Cutoff[c]	Meaning
EC_{50} (μM)[a]	≥ 168.45	Effective concentration of the nanoparticle which inhibits at 50% the growth of the assay organism.
IC_{50} (μM)[a]	≥ 177.62	Concentration of the nanoparticle which inhibits the root elongation of the assay organism (plants) at 50%.
TC_{50} (μM)[b]	≥ 225.81	Concentration which causes toxic effects in 50% of the assay organisms.
LC_{50} (μM)[b]	≥ 334.08	Lethal concentration which causes mortality in 50% of the assay organisms.

[a] Based on tests with bacteria, algae, and fungi; [b] Based on tests with crustaceans and fish; [c] Condition under which a nanoparticle was assigned to the category of non-ecotoxic.

In Equation 3, the ecotoxicity $ETOX_i\left(c_j\right)_{nw}$ of any NP in the new/final state can be obtained from the corresponding ecotoxicity $ETOX_i\left(c_j\right)_{rf}$ of another NP selected as reference, and the perturbations terms $DDX_i\left(c_j\right)$. In addition, *f* is a very complex function used to model the dependence of $ETOX_i\left(c_j\right)_{nw}$ with respect to $ETOX_i\left(c_j\right)_{rf}$ and $DDX_i\left(c_j\right)$. Notice that both $ETOX_i\left(c_j\right)_{nw}$ and $ETOX_i\left(c_j\right)_{rf}$ are binary variables, and the terms $DDX_i\left(c_j\right)$ are calculated according to the following formula:

$$DDX_i(c_j) = DX_i(c_j)_{nw} - DX_i(c_j)_{rf} \tag{4}$$

Thus, $DDX_i\left(c_j\right)$ describes the differences between any two NPs (new and references states) in terms of their variations in chemical composition, size, shape, conditions under which the sizes are measured, biological target that is tested, measures of ecotoxicity, and times of exposure of the biological targets to the NPs.

The dataset containing the 5939 cases (pairs of NPs) was randomly split into training and prediction sets. The training set was used to seek for the best QSAR-perturbation model and being formed by 4439 cases, in which 2081 cases correspond to NPs in the new state considered as non-ecotoxic and 2322 cases assigned ecotoxic. As to the prediction set, it contained 1500 cases with 715 cases considered as non-ecotoxic and 785 ecotoxic, and it was used to validate the model. The function *f* of high complexity appearing in Equation 3 was modeled by using different types of ANN. The program employed to perform this task was STATISTICA 6.0 (Statsoft-Team, 2001). In order to select the most appropriate descriptors to build up the QSAR-perturbation model, we performed a sensitivity analysis. In this sense, the neural network module of STATISTICA defined a missing value substitution procedure, which is used to allow predictions to be made in the absence of values for one or more input variables (Statsoft-Team, 2001). Thus, to define the sensitivity of a particular input variable (descriptor) *v*, in a first step, each ANN is run on a defined set of cases (training cases), where a network error is accumulated (Hill & Lewicki, 2006). Then, in a second step, the network is run again using the same cases, but this time replacing the observed values of *v* with the value estimated by the missing value procedure. So, a new network error is accumulated. Taking into account that some information that each network uses, has effectively been removed (*i.e.* one of the input variables), it is logical to expect some error' deterioration to occur (Hill & Lewicki, 2006). Then, the sensitivity of any input variable is calculated as the ratio of the error with

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

missing value substitution to the original error. The more sensitive the network is to a particular input variable (descriptor), the greater the deterioration one can expect, and therefore the greater the ratio. The larger the value of sensitivity, the greater the importance/influence a descriptor has in the model. For this reason, only variables with the largest values of sensitivity were selected. Also, we attempted to avoid collinearity between the descriptors. The statistical quality of the QSAR-perturbation model was assessed by examining some statistical indices such as the percentages of correct classifications for both non-ecotoxic and ecotoxic cases, the overall correct classification in the whole dataset, the Matthews correlation coefficient (MCC) (González-Díaz et al., 2007), and the areas under the receiver operating characteristic (ROC) curves for the training and prediction sets (Hanczar et al., 2010).

In order to search for the best QSAR-perturbation model, we tested different kinds of architectures for the ANN, namely: linear neural networks (LNN), probabilistic neural networks (PNN), radial basis function (RBF), and multi-layer perceptron (MLP). The ANN with the highest quality identified had the profile MLP 8:8-12-1:1. The symbols and meanings of the descriptors of the final QSAR-perturbation model are shown in Table 2.

In the training set, our model correctly classified 2036 of 2081 (97.84%) cases in the category of non-ecotoxic, and 2322 of 2358 (98.47%) cases in the ecotoxic category. At the same time, the results in the prediction set showed that the model rightly classified 97.20% (695 of 715) of cases assigned as non-ecotoxic, and 97.45% (765 of 785) of cases considered as ecotoxic. The accuracies displayed by this QSAR-perturbation model were 98.18% and 97.33% for the training and prediction sets, respectively. The results of the classifications can be found in the Supporting Information (**SM1** file). On the other hand, we calculated the MCC values since these measure the strength in the correlations between the observed and predicted values of categorical variables. Thus, MCC values of 0.963 and 0.947 were found for the training and prediction sets, respectively. Accordingly, in both cases there was a very strong linear correlation between the observed and predicted values of the categorical variable $ETOX_i(c_j)_{nw}$, confirming the quality and predictive power of our model.

Table 2. Descriptors that entered into the final QSAR-perturbation model

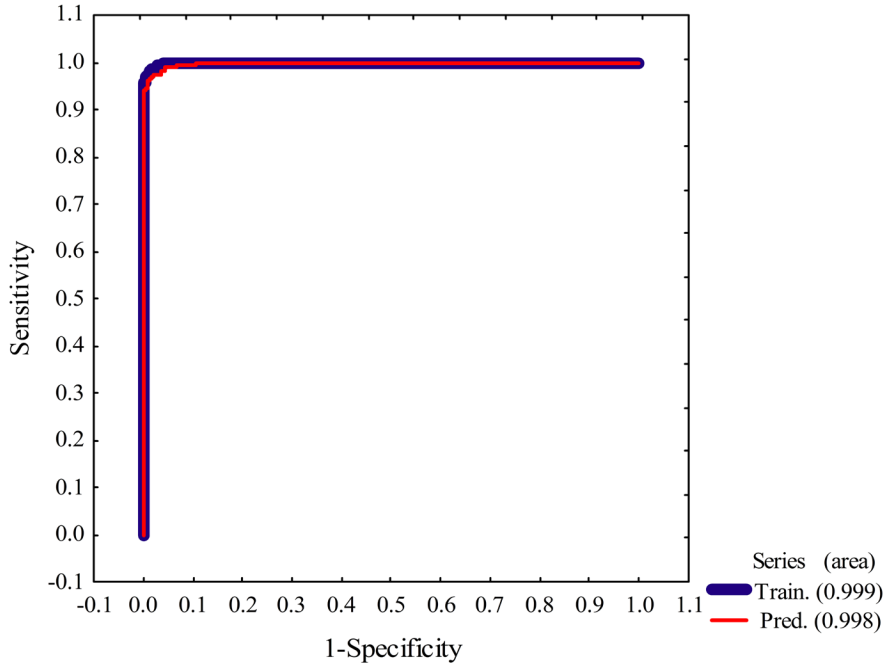
Descriptor	Meaning
$\boldsymbol{ETOX_i(c_j)_{rf}}$	Categorical variable reflecting the ecotoxicity of the nanoparticle used as reference.
$\boldsymbol{DDV(m_e)}$	Perturbation term which accounts for the changes in the molar volume between the new (output or final state) nanoparticle and that used as reference, also depending on the measures of ecotoxicity.
$\boldsymbol{DDV(b_t)}$	Perturbation term based on the variations in the molar volume between the new (output or final state) nanoparticle and that used as reference, also depending on the biological targets.
$\boldsymbol{DDE(b_t)}$	Perturbation term focused on the changes in the electronegativity between the new (output or final state) nanoparticle and that used as reference, also depending on the biological targets.
$\boldsymbol{DDP(p_s)}$	Perturbation term which describes the changes in the polarizability between the new (output or final state) nanoparticle and that used as reference, also depending on the shapes of the nanoparticles.
$\boldsymbol{DDE(d_m)}$	Perturbation term which characterizes the variations in the electronegativity between the new (output or final state) nanoparticle and that used as reference, also depending on the conditions under which the sizes of the nanoparticles were measured.
$\boldsymbol{DDL(d_m)}$	Perturbation term which takes into consideration the changes in the size between the new (output or final state) nanoparticle and that used as reference, also depending on the condition under which the sizes of the nanoparticles were measured.
$\boldsymbol{DDE(t_e)}$	Perturbation term which describes the variations in the electronegativity between the new (output or final state) nanoparticle and that used as reference, also depending on the times during which the biological targets were exposed to nanoparticles.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Finally, the areas under the ROC curves (see Figure 1) served as a final check-up of the very good performance of the derived QSAR-perturbation model. The values found were 0.999 and 0.998 for the training and prediction sets, respectively, indicating that the present model is not a random classifier because the areas under the ROC curves are very different from 0.5 – *i.e.* the value for a random classifier. The indices examined here demonstrate in terms of statistical quality and predictive power that this QSAR-perturbation model is comparable with those reported in the literature (Concu et al., 2009a; Concu et al., 2009b; Garcia et al., 2011; Gonzalez-Diaz et al., 2013; Tenorio-Borroto et al., 2013). One should notice here that all the statistical indices reported for the QSAR-perturbation model based on ANN are very similar to those reported in the second mentioned work (Kleandrova et al., 2014). However, for the present QSAR-perturbation model a considerably larger number of pairs of NPs was generated and correctly classified. These simple but essential details confirm the idea that ANN can cover a wider "nanotoxicity space".

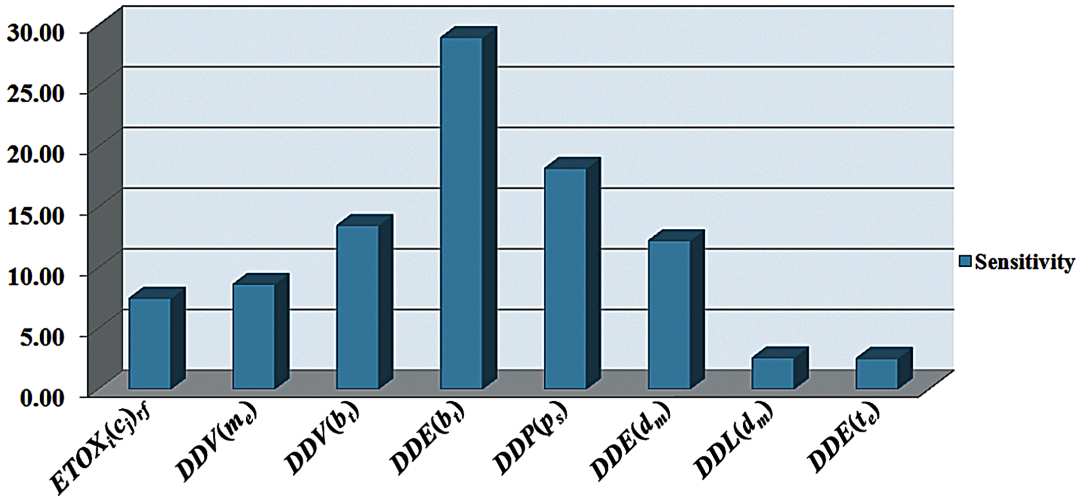
By definition, each descriptor present in any QSAR model has a degree of importance, which ensures its influence in the performance of that model. On the other hand, QSAR models should embody very useful interpretation in terms of structural and/or physicochemical properties, forward-giving an explanation regarding the changes in the activities/toxicities of the chemical species. In the case of our QSAR-perturbation model, we computed the corresponding sensitivity values of the descriptors (Figure 2). As commented before, sensitivities in ANN represent the relative importance of the variables (descriptors). Generally, for any ANN-based model built, no equation can be obtained and thus, it is very difficult to know how the descriptors should be varied to trigger particular changes in the activities/toxicities. Despite the lack of equations, one can interpret the variations of the different descriptors in such kinds of models following the procedure recently suggested by some of the present authors (Speck-Planche &

Figure 1. ROC curves obtained from the QSAR-perturbation model based on ANN



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 2. Relative influence of the descriptors in the QSAR-perturbation model



Kleandrova, 2012). That is to say, if the mean values of the perturbation terms for non-ecotoxic cases are compared to the same values for ecotoxic cases, then, the analyst can have an idea about the tendency of variation of the descriptors in the model. Hence, by analyzing Figure 2 and Table 3, we can make this kind of deductions by considering how the different factors in the NP should be modified in order to decrease its ecotoxic effect.

The descriptor ***DDE*(b_t)** has the greatest influence in the QSAR-perturbation model based on ANN. This descriptor depends on the differences between the electronegativities of the NPs used as new and reference states, and the biological targets which were tested. Inspection of Table 3 tells us that the electronegativity of the elements that constitute the NPs should diminish, *i.e.* the presence of oxygen atoms should be avoided. Diminution of ***DDE*(b_t)** will thus decrease the ecotoxic effect of the NPs. The same variation in the electronegativity characterized by ***DDE*(d_m)** (4th in raking) will produce a similar effect in the ecotoxicity. This descriptor also depends on the conditions under which the sizes of the NPs have been measured. However, these changes in the electronegativity are constrained by ***DDE*(t_e)** (having the smallest influence in the model) which additionally describes different times of exposure of the biological targets to the NPs.

Table 3. Mean values of the perturbation terms

Descriptor	Mean (Non-Ecotoxic)	Mean (Ecotoxic)
***DDV*(m_e)**	+0.742	−1.092
***DDV*(b_t)**	−0.161	+0.460
***DDE*(b_t)**	−0.213	+0.261
***DDP*(p_s)**	−0.517	+1.090
***DDE*(d_m)**	−0.154	+0.175
***DDL*(d_m)**	+9.677	−7.286
***DDE*(t_e)**	−0.043	−0.012

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

The second most important descriptor is $\boldsymbol{DDP(p_s)}$, which is based on the differences of polarizability between any two NPs, as well as in their shapes. Therefore, diminution in the polarizability of the NPs should make them less ecotoxic. With the aim of diminishing the ecotoxic effect of NPs, the values of $\boldsymbol{DDV(b_t)}$ (3rd in raking) are also to be decreased. This descriptor has a similar interpretation as $\boldsymbol{\Delta\Delta E(b_t)}$, with the differences between NPs involving molar volume instead of electronegativity. In any case, $\boldsymbol{DDV(b_t)}$ is remarkably constrained by $\boldsymbol{DDV(m_e)}$ (5th in raking) that depends on the diverse measures of ecotoxicity with which the NPs have been assessed. Notice that the combined interpretation of $\boldsymbol{DDV(b_t)}$ and $\boldsymbol{DDV(m_e)}$ indicates that there should be an equilibrium between the ability of the NPs to penetrate the cells of the biological targets and the possibility to undergo agglomeration. By considering the information provided in Table 3 for the case of the descriptor $\boldsymbol{DDL(d_m)}$ (7th in raking), one sees that the increment in the size of the NPs will diminish its potential ecotoxicity. This physical variable will depend on the difference in size of the NPs, and the conditions under which such sizes have been measured. As final aspect, we would like to emphasize that our approach depicted in Table 3 strongly converge to the physicochemical interpretation of the descriptors discussed in the previous report (Kleandrova et al., 2014).

Virtual Screening Using the QSAR-Perturbation Model

Even when any QSAR model is built on, its applicability should be demonstrated. Although the classical variants of such models restrict their use to try to understand the behavior of the chemical species, a more promising application can be realized, *i.e.* to perform virtual screening. For the specific case of our QSAR-perturbation model, we were able to give very simple physicochemical explanations about the characteristics of the NPs that can be desirable and/or detrimental to lessen the ecotoxicity. Following on, we will use our model to predict the ecotoxicity of different NPs extracted from external sources.

The first case of study is focused onto the prediction of ecotoxic effects of nickel (Ni) NPs, which have been reported experimentally a couple of years ago (Ispas et al., 2009). In such work, three different spherical Ni-NPs with sizes 30 nm, 60 nm, and 100nm (dry powders) were tested against zebra fish (*Danio rerio*) embryos during a period of 120 hrs. The acute toxicities (LC_{50}) were measured, exhibiting value of 328 mg/L (5588.40 μM), 361 mg/L (6150.65 μM), and 221 mg/L (3765.36 μM) for the sizes 30 nm, 60 nm, and 100 nm, respectively. The second case of study is devoted to predict the ecotoxicity of a CuO-NP with size 30 nm, which has been experimentally evaluated against *Pseudokirchneriella subcapitata* (Aruoja et al., 2009). This microalga was exposed to the CuO-NP (dry powder) during a period of 72 hrs, and the effective concentration leading to 50% reduction in the growth (EC_{50}) had a value of 0.71 mg Cu/L, which represents 0.889 mg/L (11.18 μM) for the CuO-NP. If we compare all these ecotoxicities with the corresponding cut off values, then, Ni-NPs should be classified as non-ecotoxic by our QSAR-perturbation model based on ANN, while the CuO-NP should be predicted as ecotoxic.

We need to emphasize that the decision of considering these new NPs as non-ecotoxic or ecotoxic by using our QSAR-perturbation model was not based on one single prediction. All the NPs were predicted 85 times, because the 85 NPs belonging to the original dataset were used as reference NPs. Accordingly, as commented in the previous section, the decision of considering NPs as non-ecotoxic (for the case of Ni-NPs) or ecotoxic (for the case of the CuO-NP) depends on the percentages of correct classification extracted from a consensus prediction, according to certain rules.

The particular results of predictions for each NP can be found in the Supporting Information (file **SM2**). Nevertheless, the summary of the results of the predictions are depicted in Table 4. The spherical Ni-NPs with sizes 30 nm, 60 nm, and 100 nm were correctly predicted as non-ecotoxic with very high

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

consensus percentages. Of particular interest are the differences in the consensus percentages between the three Ni-NPs. Notice that the only condition which varies is the size of each NP, and the larger the size the higher the consensus percentage indicating the low ecotoxicity of the Ni-NP. These results have a strong convergence with our explanation regarding the physicochemical interpretations of the perturbation terms, where we suggested that the size of any NP should be increased in order to decrease its ecotoxic effect. On the other hand, predictions performed for the CuO-NP were also in agreement with the experimental evidence. In this case, the authors of the research did not report the shape of the CuO-NP (Aruoja et al., 2009). Therefore, we decided to predict the ecotoxicity of the CuO-NP by considering six of the shape labels understudy, namely: spherical, pyramidal, irregular, elliptical, needle, and strip. We also took into account the category where no NP shape has been reported (N/A). Our predictions indicate us that CuO-NP is toxic against *Pseudokirchneriella subcapitata*, for all the possible shapes considered by the QSAR-perturbation model (Table 4). Therefore, the most important factor associated with the ecotoxicity of the CuO-NP is not the shape but its composition. That comes from the fact that the CuO-NP has one very electronegative element (oxygen), and as commented above, the presence of electronegative atoms can increase the ecotoxic effect of NPs. Both these suggestions and the prediction results demonstrate that our QSAR-perturbation model is very sensitive to the changes in both physicochemical properties associated with the composition, and different experimental elements, being thus a promising computational tool for the efficient prediction of ecotoxicity of NPs under diverse sets of experimental conditions.

Table 4. Consensus predictions performed using the QSAR-perturbation model

Nanoparticle	m_e	b_t	p_s	d_m	t_e	$ETOX_i(c_j)_{nw}$	%CCT[a]
Ni-30 nm	LC_{50} (μM)	*Danio rerio* (embryos)	spherical	dry	120	+1	84.71
Ni-60 nm	LC_{50} (μM)	*Danio rerio* (embryos)	spherical	dry	120	+1	89.41
Ni-100 nm	LC_{50} (μM)	*Danio rerio* (embryos)	spherical	dry	120	+1	96.47
CuO-30 nm	EC_{50} (μM)	*Pseudokirchneriella subcapitata*	N/A	dry	72	−1	87.06
CuO-30 nm	EC_{50} (μM)	*Pseudokirchneriella subcapitata*	spherical	dry	72	−1	97.65
CuO-30 nm	EC_{50} (μM)	*Pseudokirchneriella subcapitata*	pyramidal	dry	72	−1	76.47
CuO-30 nm	EC_{50} (μM)	*Pseudokirchneriella subcapitata*	irregular	dry	72	−1	76.47
CuO-30 nm	EC_{50} (μM)	*Pseudokirchneriella subcapitata*	elliptical	dry	72	−1	65.88
CuO-30 nm	EC_{50} (μM)	*Pseudokirchneriella subcapitata*	needle	dry	72	−1	57.65
CuO-30 nm	EC_{50} (μM)	*Pseudokirchneriella subcapitata*	strip	dry	72	−1	100.00

[a] Consensus percentage, *i.e.* the percentage of times that the QSAR-perturbation model correctly predicted the nanoparticle according to its observed value of $ETOX_i(c_j)_{nw}$.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

CONCLUDING REMARKS AND RECOMMENDATIONS

Several researchers have emphasized the use of promising guidelines for the computational modeling of toxic effects of nanomaterials (Kahru & Ivask, 2013; Westerhoff & Nowack, 2013), and there are also some reviews and chapters directed towards the development of reliable nano-QSAR models in recent years (Fourches et al., 2010; Kar et al., 2014a; Puzyn et al., 2011; Toropov et al., 2012). It is genuinely true that such task is a great challenge that may lead to new and exciting directions for QSAR modeling techniques. However, there are some problems one should face, namely:

- In the process of building reliable QSAR models, the ecotoxicological profiles of nanomaterials should be clear along with the development of toxicology studies.
- There is a strong need to supplement the existing set of molecular descriptors by novel "nano-descriptors" that can represent size, shape, and so on-dependent properties of nanomaterials.
- More sophisticated methods such as the present QSAR-perturbation theory should be followed whenever developing QSAR models to tackle nanomaterials' endpoints.
- Most likely the current criteria of evaluating the models' quality (*i.e.* the five OECD rules) will have to be re-evaluated and adapted to nanomaterials. This is due to the specific properties of chemicals occurring at the "nano" level (for instance, electronic properties change with changing size), and the very limited number of available data (Notice that one has to overcome the problems of the "classical" method of validation that is biased towards small, low molecular weight molecules.).
- Regarding composition, nanoparticles in biological media are covered by a stabilizing layer and/or a protein corona, so that the NP surface may not be in direct contact with cells, so a change in the toxicity profile is to be expected and should be taken into account.

FUTURE RESEARCH DIRECTIONS

In this chapter, we have developed the first unified QSAR-perturbation model, demonstrating its ability to predict ecotoxicity of NPs under diverse experimental conditions. In so doing, the next promising horizons regarding the application of QSAR-perturbation models are to be devoted to the integration of different ecotoxicity profiles, with other biological effects such as cytotoxicities, and pharmacological activities (antibacterial, antiviral, anticancer). This will ensure the design and application of safer nanoparticles in very important fields like nanomedicine. On the other hand, further studies aiming at deriving QSAR-perturbation models to predict synthesis of NPs can contribute to the creation of more sustainable processes and/or alternatives in nanochemistry, helping to control the release of NPs into the environment. Finally, we would like to point out that the use of complex network theory may well be of great interest in order to gain a more thorough knowledge about the ecotoxic effects of NPs. Towards that goal, the design and study of the dynamics of NP-NP complex networks may help to understand the evolution of the ecotoxic behavior of NPs over time. All these ideas will contribute towards further developing and establishing one discipline that has just emerged as a need to study the 'nanoworld' in a more rational way: *nanoinformatics*.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

ACKNOWLEDGMENT

The authors thank the State Key Laboratory of Environmental Chemistry and Ecotoxicology, Research Centre for Eco-Environmental Sciences, Chinese Academy of Sciences for financial support (KF2013-09). This work is also supported by Grant No. Pest-C/EQB/LA0006/2013 and Project NORTE-07-0124-FEDER-000067-NANOCHEMISTRY financed by the Portuguese *Fundação para a Ciência e a Tecnologia* (FCT) and FEDER. FL and ASP acknowledge the Grants SFRH/BPD/63666/2009 and SFRH/BD/77690/2011, respectively, financed by FCT and the European Social Fund.

REFERENCES

Alivisatos, P. (2004). The use of nanocrystals in biological detection. *Nature Biotechnology*, *22*(1), 47–52. doi:10.1038/nbt927 PMID:14704706

Allen, C., Maysinger, D., & Eisenberg, A. (1999). Nano-engineering block copolymer aggregates for drug delivery. *Colloids and Surfaces. B, Biointerfaces*, *16*(1-4), 3–27. doi:10.1016/S0927-7765(99)00058-2

Aruoja, V., Dubourguier, H. C., Kasemets, K., & Kahru, A. (2009). Toxicity of nanoparticles of CuO, ZnO and TiO_2 to microalgae Pseudokirchneriella subcapitata. *The Science of the Total Environment*, *407*(4), 1461–1468. doi:10.1016/j.scitotenv.2008.10.053 PMID:19038417

Bakry, R., Vallant, R. M., Najam-ul-Haq, M., Rainer, M., Szabo, Z., Huck, C. W., & Bonn, G. K. (2007). Medicinal applications of fullerenes. *International Journal of Nanomedicine*, *2*(4), 639–649. PMID:18203430

Bar-Ilan, O., Albrecht, R. M., Fako, V. E., & Furgeson, D. Y. (2009). Toxicity assessments of multisized gold and silver nanoparticles in zebrafish embryos. *Small*, *5*(16), 1897–1910. doi:10.1002/smll.200801716 PMID:19437466

Bianco, A., Kostarelos, K., Partidos, C. D., & Prato, M. (2005). Biomedical applications of functionalised carbon nanotubes. *Chemical Communications*, *5*(5), 571–577. doi:10.1039/b410943k PMID:15672140

Borm, P. J., Robbins, D., Haubold, S., Kuhlbusch, T., Fissan, H., & Donaldson, K. et al. (2006). The potential risks of nanomaterials: A review carried out for ECETOC. *Particle and Fibre Toxicology*, *3*(1), 11. doi:10.1186/1743-8977-3-11 PMID:16907977

Bréchignac, C., Houdy, P., & Lahmani, M. (Eds.). (2007). Nanomaterials and nanochemistry. Berlin: Springer Science+Bussiness Media. doi:10.1007/978-3-540-72993-8

Cai, R., Van, G. M., Aw, P. K., & Itoh, K. (2006). Solar-driven self-cleaning coating for a painted surface. *Comptes Rendus. Chimie*, *9*(5-6), 829–835. doi:10.1016/j.crci.2005.04.007

Chen, F. Q., & Gerion, D. (2004). Fluorescent CdSe/ZnS nanocrystal-peptide conjugates for long-term, nontoxic imaging and nuclear targeting in living cells. *Nano Letters*, *4*(10), 1827–1832. doi:10.1021/nl049170q

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Concu, R., Dea-Ayuela, M. A., Perez-Montoto, L. G., Prado-Prado, F. J., & Uriarte, E. (2009). 3D entropy and moments prediction of enzyme classes and experimental-theoretic study of peptide fingerprints in Leishmania parasites. *Biochimica et Biophysica Acta*, (12), 1784–1794. PMID:19716935

Concu, R., Podda, G., Uriarte, E., & Gonzalez-Diaz, H. (2009). Computational chemistry study of 3D-structure-function relationships for enzymes based on Markov models for protein electrostatic, HINT, and van der Waals potentials. *Journal of Computational Chemistry*, *30*(9), 1510–1520. doi:10.1002/jcc.21170 PMID:19086060

Cruz-Monteagudo, M., Ancede-Gallardo, E., Jorge, M., & Cordeiro, M. N. D. S. (2013). Chemoinformatics profiling of ionic liquids--automatic and chemically interpretable cytotoxicity profiling, virtual screening, and cytotoxicophore identification. *Toxicological Sciences*, *136*(2), 548–565. doi:10.1093/toxsci/kft209 PMID:24068674

Cruz-Monteagudo, M., Gonzalez-Diaz, H., Borges, F., Dominguez, E. R., & Cordeiro, M. N. D. S. (2008). 3D-MEDNEs: An alternative "in silico" technique for chemical research in toxicology. 2. quantitative proteome-toxicity relationships (QPTR) based on mass spectrum spiral entropy. *Chemical Research in Toxicology*, *21*(3), 619–632. doi:10.1021/tx700296t PMID:18257557

Cruz-Monteagudo, M., Gonzalez-Diaz, H., Borges, F., & Gonzalez-Diaz, Y. (2006). Simple stochastic fingerprints towards mathematical modeling in biology and medicine. 3. Ocular irritability classification model. *Bulletin of Mathematical Biology*, *68*(7), 1555–1572. doi:10.1007/s11538-006-9083-y PMID:16865609

De Volder, M. F., Tawfick, S. H., Baughman, R. H., & Hart, A. J. (2013). Carbon nanotubes: Present and future commercial applications. *Science*, *339*(6119), 535–539. doi:10.1126/science.1222453 PMID:23372006

Dimitrov, S., Dimitrova, G., Pavlov, T., Dimitrova, N., Patlewicz, G., Niemela, J., & Mekenyan, O. (2005). A stepwise approach for defining the applicability domain of SAR and QSAR models. *Journal of Chemical Information and Modeling*, *45*(4), 839–849. doi:10.1021/ci0500381 PMID:16045276

Dragos, H., Gilles, M., & Alexandre, V. (2009). Predicting the predictability: A unified approach to the applicability domain problem of QSAR models. *Journal of Chemical Information and Modeling*, *49*(7), 1762–1776. doi:10.1021/ci9000579 PMID:19530661

Drexler, K. E. (1992). Nanotechnology: The past and the future. *Science*, *255*(5042), 268–269. doi:10.1126/science.255.5042.268-b PMID:17779552

Eisenbrand, G., Pool-Zobel, B., Baker, V., Balls, M., Blaauboer, B. J., & Boobis, A. et al. (2002). Methods of in vitro toxicology. *Food and Chemical Toxicology*, *40*(2-3), 193–236. doi:10.1016/S0278-6915(01)00118-1 PMID:11893398

Estrada, E., Molina, E., & Uriarte, E. (2001). Quantitative structure-toxicity relationships using TOPS-MODE. 2. Neurotoxicity of a non-congeneric series of solvents. *SAR and QSAR in Environmental Research*, *12*(5), 445–459. doi:10.1080/10629360108035384 PMID:11813810

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Estrada, E., & Peña, A. (2000). In silico studies for the rational discovery of anticonvulsant compounds. *Bioorganic & Medicinal Chemistry*, *8*(12), 2755–2770. doi:10.1016/S0968-0896(00)00204-2 PMID:11131167

Estrada, E., Pena, A., & Garcia-Domenech, R. (1998). Designing sedative/hypnotic compounds from a novel substructural graph-theoretical approach. *Journal of Computer-Aided Molecular Design*, *12*(6), 583–595. doi:10.1023/A:1008048003720 PMID:9879506

Estrada, E., Quincoces, J. A., & Patlewicz, G. (2004). Creating molecular diversity from antioxidants in Brazilian propolis. Combination of TOPS-MODE QSAR and virtual structure generation. *Molecular Diversity*, *8*(1), 21–33. doi:10.1023/B:MODI.0000006804.97390.40 PMID:14964785

Estrada, E., & Uriarte, E. (2001). Quantitative structure-toxicity relationships using TOPS-MODE. 1. Nitrobenzene toxicity to Tetrahymena pyriformis. *SAR and QSAR in Environmental Research*, *12*(3), 309–324. doi:10.1080/10629360108032919 PMID:11696927

Estrada, E., Uriarte, E., Gutierrez, Y., & Gonzalez, H. (2003). Quantitative structure-toxicity relationships using TOPS-MODE. 3. Structural factors influencing the permeability of commercial solvents through living human skin. *SAR and QSAR in Environmental Research*, *14*(2), 145–163. doi:10.1080/1062936031000073162 PMID:12747572

Estrada, E., Uriarte, E., Molina, E., Simon-Manso, Y., & Milne, G. W. (2006). An integrated in silico analysis of drug-binding to human serum albumin. *Journal of Chemical Information and Modeling*, *46*(6), 2709–2724. doi:10.1021/ci600274f PMID:17125211

Estrada, E., Uriarte, E., Montero, A., Teijeira, M., Santana, L., & De Clercq, E. (2000). A novel approach for the virtual screening and rational design of anticancer compounds. *Journal of Medicinal Chemistry*, *43*(10), 1975–1985. doi:10.1021/jm991172d PMID:10821710

Fechner, N., Jahn, A., Hinselmann, G., & Zell, A. (2010). Estimation of the applicability domain of kernel-based machine learning models for virtual screening. *Journal of Cheminformatics*, *2*(1), 2. doi:10.1186/1758-2946-2-2 PMID:20222949

Feynman, R. P. (1960). There's plenty of room at the bottom (an invitation to enter a new field of physics). *Engineering and Science*, *23*(5), 22–36.

Fourches, D., Pu, D., Tassa, C., Weissleder, R., Shaw, S. Y., Mumper, R. J., & Tropsha, A. (2010). Quantitative nanostructure-activity relationship modeling. *ACS Nano*, *4*(10), 5703–5712. doi:10.1021/nn1013484 PMID:20857979

García, A., Espinosa, R., Delgado, L., Casals, E., González, E., & Puntes, V. et al. (2011). Acute toxicity of cerium oxide, titanium oxide and iron oxide nanoparticles using standardized tests. *Desalination*, *269*(1-3), 136–141. doi:10.1016/j.desal.2010.10.052

Garcia, I., Fall, Y., Gomez, G., & Gonzalez-Diaz, H. (2011). First computational chemistry multi-target model for anti-Alzheimer, anti-parasitic, anti-fungi, and anti-bacterial activity of GSK-3 inhibitors in vitro, in vivo, and in different cellular lines. *Molecular Diversity*, *15*(2), 561–567. doi:10.1007/s11030-010-9280-3 PMID:20931280

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Gong, N., Shao, K., Feng, W., Lin, Z., Liang, C., & Sun, Y. (2011). Biotoxicity of nickel oxide nanoparticles and bio-remediation by microalgae Chlorella vulgaris. *Chemosphere*, *83*(4), 510–516. doi:10.1016/j.chemosphere.2010.12.059 PMID:21216429

Gonzalez-Diaz, H., Arrasate, S., Gomez-Sanjuan, A., Sotomayor, N., Lete, E., Besada-Porto, L., & Ruso, J. M. (2013). General theory for multiple input-output perturbations in complex molecular systems. 1. Linear QSPR electronegativity models in physical, organic, and medicinal chemistry. *Current Topics in Medicinal Chemistry*, *13*(14), 1713–1741. doi:10.2174/1568026611313140011 PMID:23889050

Gonzalez-Diaz, H., Herrera-Ibata, D. M., Duardo-Sanchez, A., Munteanu, C. R., Orbegozo-Medina, R. A., & Pazos, A. (2014). ANN multiscale model of anti-HIV drugs activity vs AIDS prevalence in the US at county level based on information indices of molecular graphs and social networks. *Journal of Chemical Information and Modeling*, *54*(3), 744–755. doi:10.1021/ci400716y PMID:24521170

Gonzalez-Díaz, H., Perez-Bello, A., Cruz-Monteagudo, M., Gonzalez-Diaz, Y., Santana, L., & Uriarte, E. (2007). Chemometrics for QSAR with low sequence homology: Mycobacterial promoter sequences recognition with 2D-RNA entropies. *Chemometrics and Intelligent Laboratory Systems*, *85*(1), 20–26. doi:10.1016/j.chemolab.2006.03.005

Griffitt, R. J., Luo, J., Gao, J., Bonzongo, J. C., & Barber, D. S. (2008). Effects of particle composition and species on toxicity of metallic nanomaterials in aquatic organisms. *Environmental Toxicology and Chemistry*, *27*(9), 1972–1978. doi:10.1897/08-002.1 PMID:18690762

Hanczar, B., Hua, J., Sima, C., Weinstein, J., Bittner, M., & Dougherty, E. R. (2010). Small-sample precision of ROC-related estimates. *Bioinformatics (Oxford, England)*, *26*(6), 822–830. doi:10.1093/bioinformatics/btq037 PMID:20130029

Hansch, C., & Leo, A. (1995). *Exploring QSAR: fundamentals and applications in chemistry and biology*. Washington, DC: American Chemical Society.

Heinlaan, M., Ivask, A., Blinova, I., Dubourguier, H. C., & Kahru, A. (2008). Toxicity of nanosized and bulk ZnO, CuO and TiO_2 to bacteria Vibrio fischeri and crustaceans Daphnia magna and Thamnocephalus platyurus. *Chemosphere*, *71*(7), 1308–1316. doi:10.1016/j.chemosphere.2007.11.047 PMID:18194809

Helguera, A. M., Cabrera Perez, M. A., Gonzalez, M. P., Ruiz, R. M., & Gonzalez Diaz, H. (2005). A topological substructural approach applied to the computational prediction of rodent carcinogenicity. *Bioorganic & Medicinal Chemistry*, *13*(7), 2477–2488. doi:10.1016/j.bmc.2005.01.035 PMID:15755650

Helguera, A. M., Gonzalez, M. P., Cordeiro, M. N. D. S., & Perez, M. A. (2007). Quantitative structure carcinogenicity relationship for detecting structural alerts in nitroso-compounds. *Toxicology and Applied Pharmacology*, *221*(2), 189–202. doi:10.1016/j.taap.2007.02.021 PMID:17477948

Hill, T., & Lewicki, P. (2006). *STATISTICS methods and applications. A comprehensive reference for science, industry and data mining*. Tulsa, OK: StatSoft.

Holden, P. A., Nisbet, R. M., Lenihan, H. S., Miller, R. J., Cherr, G. N., Schimel, J. P., & Gardea-Torresdey, J. L. (2013). Ecological nanotoxicology: Integrating nanomaterial hazard considerations across the subcellular, population, community, and ecosystems levels. *Accounts of Chemical Research*, *46*(3), 813–822. doi:10.1021/ar300069t PMID:23039211

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Hsu, D. D. (2013). *Chemicool periodic table*. Retrieved from http://www.chemicool.com/

Hund-Rinke, K., & Simon, M. (2006). Ecotoxic effect of photocatalytic active nanoparticles (TiO2) on algae and daphnids. *Environmental Science and Pollution Research International*, *13*(4), 225–232. doi:10.1065/espr2006.06.311 PMID:16910119

Ispas, C., Andreescu, D., Patel, A., Goia, D. V., Andreescu, S., & Wallace, K. N. (2009). Toxicity and developmental defects of different sizes and shape nickel nanoparticles in zebrafish. *Environmental Science & Technology*, *43*(16), 6349–6356. doi:10.1021/es9010543 PMID:19746736

Kahru, A., & Ivask, A. (2013). Mapping the dawn of nanoecotoxicological research. *Accounts of Chemical Research*, *46*(3), 823–833. doi:10.1021/ar3000212 PMID:23148404

Kar, S., Gajewicz, A., Puzyn, T., & Roy, K. (2014a). Nano-quantitative structure-activity relationship modeling using easily computable and interpretable descriptors for uptake of magnetofluorescent engineered nanoparticles in pancreatic cancer cells. *Toxicology in Vitro*, *28*(4), 600–606. doi:10.1016/j.tiv.2013.12.018 PMID:24412539

Kar, S., Gajewicz, A., Puzyn, T., Roy, K., & Leszczynski, J. (2014b). Periodic table-based descriptors to encode cytotoxicity profile of metal oxide nanoparticles: A mechanistic QSTR approach. *Ecotoxicology and Environmental Safety*, *107*, 162–169. doi:10.1016/j.ecoenv.2014.05.026 PMID:24949897

Kasemets, K., Ivask, A., Dubourguier, H. C., & Kahru, A. (2009). Toxicity of nanoparticles of ZnO, CuO and TiO_2 to yeast Saccharomyces cerevisiae. *Toxicology In Vitro*, *23*(6), 1116–1122. doi:10.1016/j.tiv.2009.05.015 PMID:19486936

Kim, Y., Dalhaimer, P., Christian, D. A., & Discher, D. E. (2005). Polymeric worm micelles as nano-carriers for drug delivery. *Nanotechnology*, *16*(7), S484–S491. doi:10.1088/0957-4484/16/7/024 PMID:21727469

Kleandrova, V. V., Luan, F., Gonzalez-Diaz, H., Ruso, J. M., Melo, A., Speck-Planche, A., & Cordeiro, M. N. (2014). Computational ecotoxicology: Simultaneous prediction of ecotoxic effects of nanoparticles under different experimental conditions. *Environment International*, *73*, 288–294. doi:10.1016/j.envint.2014.08.009 PMID:25173945

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

Kulkarni, S. A., & Zhu, J. (2008). Integrated approach to assess the domain of applicability of some commercial (Q)SAR models. *SAR and QSAR in Environmental Research*, *19*(1-2), 39–54. doi:10.1080/10629360701843508 PMID:18311633

Lanone, S., Rogerieux, F., Geys, J., Dupont, A., Maillot-Marechal, E., & Boczkowski, J. et al. (2009). Comparative toxicity of 24 manufactured nanoparticles in human alveolar epithelial and macrophage cell lines. *Particle and Fibre Toxicology*, *6*(1), 14. doi:10.1186/1743-8977-6-14 PMID:19405955

Lin, D., & Xing, B. (2007). Phytotoxicity of nanoparticles: Inhibition of seed germination and root growth. *Environmental Pollution*, *150*(2), 243–250. doi:10.1016/j.envpol.2007.01.016 PMID:17374428

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Liu, R., Rallo, R., George, S., Ji, Z., Nair, S., Nel, A. E., & Cohen, Y. (2011). Classification Nano-SAR development for cytotoxicity of metal oxide nanoparticles. *Small*, *7*(8), 1118–1126. doi:10.1002/smll.201002366 PMID:21456088

Liu, R., Rallo, R., Weissleder, R., Tassa, C., Shaw, S., & Cohen, Y. (2013). Nano-SAR development for bioactivity of nanoparticles with considerations of decision boundaries. *Small*, *9*(9-10), 1842–1852. doi:10.1002/smll.201201903 PMID:23423856

Luan, F., Cordeiro, M. N. D. S., Alonso, N., Garcia-Mera, X., Caamano, O., & Romero-Duran, F. J. et al. (2013). TOPS-MODE model of multiplexing neuroprotective effects of drugs and experimental-theoretic study of new 1,3-rasagiline derivatives potentially useful in neurodegenerative diseases. *Bioorganic & Medicinal Chemistry*, *21*(7), 1870–1879. doi:10.1016/j.bmc.2013.01.035 PMID:23415089

Luan, F., Kleandrova, V. V., Gonzalez-Diaz, H., Ruso, J. M., Melo, A., Speck-Planche, A., & Cordeiro, M. N. D. S. (2014). Computer-aided nanotoxicology: Assessing cytotoxicity of nanoparticles under diverse experimental conditions by using a novel QSTR-perturbation approach. *Nanoscale*, *6*(18), 10623–10630. doi:10.1039/C4NR01285B PMID:25083742

Ma, Y., He, X., Zhang, P., Zhang, Z., Guo, Z., & Tai, R. et al. (2011). Phytotoxicity and biotransformation of La_2O_3 nanoparticles in a terrestrial plant cucumber (Cucumis sativus). *Nanotoxicology*, *5*(4), 743–753. doi:10.3109/17435390.2010.545487 PMID:21261455

Ma, Y., Kuang, L., He, X., Bai, W., Ding, Y., & Zhang, Z. et al. (2010). Effects of rare earth oxide nanoparticles on root elongation of plants. *Chemosphere*, *78*(3), 273–279. doi:10.1016/j.chemosphere.2009.10.050 PMID:19897228

Marsalek, B., Jancula, D., Marsalkova, E., Mashlan, M., Safarova, K., Tucek, J., & Zboril, R. (2012). Multimodal action and selective toxicity of zerovalent iron nanoparticles against cyanobacteria. *Environmental Science & Technology*, *46*(4), 2316–2323. doi:10.1021/es2031483 PMID:22242974

Morales Helguera, A., Perez Gonzalez, M., Cordeiro, M. N. D. S., & Cabrera Perez, M. A. (2008). Quantitative structure-carcinogenicity relationship for detecting structural alerts in nitroso compounds: Species, rat; sex, female; route of administration, gavage. *Chemical Research in Toxicology*, *21*(3), 633–642. doi:10.1021/tx700336n PMID:18293904

Netzeva, T. I., Worth, A., Aldenberg, T., Benigni, R., Cronin, M. T., Gramatica, P., … Yang, C. (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 Animal, 33*(2), 155-173.

Oberdorster, G. (2010). Safety assessment for nanotechnology and nanomedicine: Concepts of nanotoxicology. *Journal of Internal Medicine*, *267*(1), 89–105. doi:10.1111/j.1365-2796.2009.02187.x PMID:20059646

Oberdörster, G., Stone, V., & Donaldson, K. (2007). Toxicology of nanoparticles: A historical perspective. *Nanotoxicology*, *1*(1), 2–25. doi:10.1080/17435390701314761

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Olson, M. S., & Gurian, P. L. (2012). Risk assessment strategies as nanomaterials transition into commercial applications. *Journal of Nanoparticle Research*, *14*(4), 786. doi:10.1007/s11051-012-0786-8

Poater, A., Saliner, A. G., Carbo-Dorca, R., Poater, J., Sola, M., Cavallo, L., & Worth, A. P. (2009). Modeling the structure-property relationships of nanoneedles: A journey toward nanomedicine. *Journal of Computational Chemistry*, *30*(2), 275–284. doi:10.1002/jcc.21041 PMID:18615420

Popat, K. C., Eltgroth, M., LaTempa, T. J., Grimes, C. A., & Desai, T. A. (2007). Titania nanotubes: A novel platform for drug-eluting coatings for medical implants? *Small*, *3*(11), 1878–1881. doi:10.1002/smll.200700412 PMID:17935080

Portney, N. G., & Ozkan, M. (2006). Nano-oncology: Drug delivery, imaging, and sensing. *Analytical and Bioanalytical Chemistry*, *384*(3), 620–630. doi:10.1007/s00216-005-0247-7 PMID:16440195

Prado-Prado, F. J., Garcia, I., Garcia-Mera, X., & Gonzalez-Diaz, H. (2011). Entropy multi-target QSAR model for prediction of antiviral drug complex networks. *Chemometrics and Intelligent Laboratory Systems*, *107*(2), 227–233. doi:10.1016/j.chemolab.2011.02.003

Prado-Prado, F. J., Garcia-Mera, X., & Gonzalez-Diaz, H. (2010). Multi-target spectral moment QSAR versus ANN for antiparasitic drugs against different parasite species. *Bioorganic & Medicinal Chemistry*, *18*(6), 2225–2231. doi:10.1016/j.bmc.2010.01.068 PMID:20185316

Puzyn, T., Rasulev, B., Gajewicz, A., Hu, X., Dasari, T. P., & Michalkova, A. 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

Rai, M., Kon, K., Ingle, A., Duran, N., Galdiero, S., & Galdiero, M. (2014). Broad-spectrum bioactivities of silver nanoparticles: The emerging trends and future prospects. *Applied Microbiology and Biotechnology*, *98*(5), 1951–1961. doi:10.1007/s00253-013-5473-x PMID:24407450

Ravi Kumar, M. N. (2000). Nano and microparticles as controlled drug delivery devices. *Journal of Pharmacy & Pharmaceutical Sciences*, *3*(2), 234–258. PMID:10994037

Sadiq, I. M., Pakrashi, S., Chandrasekaran, N., & Mukherjee, A. (2011). Studies on toxicity of aluminum oxide (Al_2O_3) nanoparticles to microalgae species: *Scenedesmus sp.* and *Chlorella sp. Journal of Nanoparticle Research*, *13*(8), 3287–3299. doi:10.1007/s11051-011-0243-0

Schrand, A. M., Dai, L. M., Schlager, J. J., & Hussain, S. M. (2012). Toxicity testing of nanomaterials. In New technologies for toxicity testing (pp. 58-75). New York: Landes Bioscience and Springer Science+Bussiness Media, LLC.

Serpone, N., Dondi, D., & Albini, A. (2007). Inorganic and organic UV filters: Their role and efficiency in sunscreens and suncare products. *Inorganica Chimica Acta*, *360*(3), 794–802. doi:10.1016/j.ica.2005.12.057

Shaw, S. Y., Westly, E. C., Pittet, M. J., Subramanian, A., Schreiber, S. L., & Weissleder, R. (2008). Perturbational profiling of nanomaterial biologic activity. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(21), 7387–7392. doi:10.1073/pnas.0802878105 PMID:18492802

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Shvedova, A. A., & Kagan, V. E. (2010). The role of nanotoxicology in realizing the 'helping without harm' paradigm of nanomedicine: Lessons from studies of pulmonary effects of single-walled carbon nanotubes. *Journal of Internal Medicine*, *267*(1), 106–118. doi:10.1111/j.1365-2796.2009.02188.x PMID:20059647

Simon, F., Peterlik, H., Pfeiffer, R., Bernardi, J., & Kuzmany, H. (2007). Fullerene release from the inside of carbon nanotubes: A possible route toward drug delivery. *Chemical Physics Letters*, *445*(4-6), 288–292. doi:10.1016/j.cplett.2007.08.014

Speck-Planche, A., & Cordeiro, M. N. D. S. (2013). Simultaneous modeling of antimycobacterial activities and ADMET profiles: A chemoinformatic approach to medicinal chemistry. *Current Topics in Medicinal Chemistry*, *13*(14), 1656–1665. doi:10.2174/15680266113139990116 PMID:23889052

Speck-Planche, A., & Cordeiro, M. N. D. S. (2014). Simultaneous virtual prediction of anti-escherichia coli activities and ADMET profiles: A chemoinformatic complementary approach for high-throughput screening. *ACS Combinatorial Science*, *16*(2), 78–84. doi:10.1021/co400115s PMID:24383958

Speck-Planche, A., & Kleandrova, V. V. (2012). QSAR and molecular docking techniques for the discovery of potent monoamine oxidase B inhibitors: Computer-aided generation of new rasagiline bioisosteres. *Current Topics in Medicinal Chemistry*, *12*(16), 1734–1747. doi:10.2174/156802612803989282 PMID:23030609

Speck-Planche, A., Kleandrova, V. V., & Cordeiro, M. N. D. S. (2013a). New insights toward the discovery of antibacterial agents: Multi-tasking QSBER model for the simultaneous prediction of anti-tuberculosis activity and toxicological profiles of drugs. *European Journal of Pharmaceutical Sciences*, *48*(4-5), 812–818. doi:10.1016/j.ejps.2013.01.011 PMID:23376211

Speck-Planche, A., Kleandrova, V. V., Luan, F., & Cordeiro, M. N. D. S. (2012). Predicting multiple ecotoxicological profiles in agrochemical fungicides: A multi-species chemoinformatic approach. *Ecotoxicology and Environment Safety, 80*, 308-313. doi:10.1016/j.ecoenv.2012.03.018

Speck-Planche, A., Kleandrova, V. V., Luan, F., & Cordeiro, M. N. D. S. (2013b). Multi-target inhibitors for proteins associated with Alzheimer: In silico discovery using fragment-based descriptors. *Current Alzheimer Research*, *10*(2), 117–124. doi:10.2174/1567205011310020001 PMID:22515494

Statsoft-Team. (2001). STATISTICA. Data analysis software system. v6.0, Tulsa.

Stupp, S., Donners, J., Li, L., & Mata, A. (2005). Expanding frontiers in biomaterials. *MRS Bulletin*, *30*(11), 864–873. doi:10.1557/mrs2005.276

Taniguchi, N. (1974). On the basic concept of 'nano-technology'. Japan Society of Precision Engineering.

Tenorio-Borroto, E., Garcia-Mera, X., Penuelas-Rivas, C. G., Vasquez-Chagoyan, J. C., Prado-Prado, F. J., Castanedo, N., & Gonzalez-Diaz, H. (2013). Entropy model for multiplex drug-target interaction endpoints of drug immunotoxicity. *Current Topics in Medicinal Chemistry*, *13*(14), 1636–1649. doi:10.2174/15680266113139990114 PMID:23889053

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Tenorio-Borroto, E., Penuelas Rivas, C. G., Vasquez Chagoyan, J. C., Castanedo, N., Prado-Prado, F. J., Garcia-Mera, X., & Gonzalez-Diaz, H. (2012). ANN multiplexing model of drugs effect on macrophages; theoretical and flow cytometry study on the cytotoxicity of the anti-microbial drug G1 in spleen. *Bioorganic & Medicinal Chemistry*, *20*(20), 6181–6194. doi:10.1016/j.bmc.2012.07.020 PMID:22981917

Tenorio-Borroto, E., Penuelas-Rivas, C. G., Vasquez-Chagoyan, J. C., Castanedo, N., Prado-Prado, F. J., Garcia-Mera, X., & Gonzalez-Diaz, H. (2014). Model for high-throughput screening of drug immunotoxicity - Study of the anti-microbial G1 over peritoneal macrophages using flow cytometry. *European Journal of Medicinal Chemistry*, *72*, 206–220. doi:10.1016/j.ejmech.2013.08.035 PMID:24445280

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

Todeschini, R., & Consonni, V. (2009). *Molecular descriptors for chemoinformatics*. Weinheim: WILEY-VCH Verlag GmbH & Co. KGaA.

Toropov, A. A., Toropova, A. P., Benfenati, E., Gini, G., Puzyn, T., Leszczynska, D., & Leszczynski, J. (2012). Novel application of the CORAL software to model cytotoxicity of metal oxide nanoparticles to bacteria Escherichia coli. *Chemosphere*, *89*(9), 1098–1102. doi:10.1016/j.chemosphere.2012.05.077 PMID:22704203

van de Waterbeemd, H. (1995). *Chemometrics methods in molecular design*. Weinheim: VCH Publishers. doi:10.1002/9783527615452

Vinogradov, S. V., Bronich, T. K., & Kabanov, A. V. (2002). Nanosized cationic hydrogels for drug delivery: Preparation, properties and interactions with cells. *Advanced Drug Delivery Reviews*, *54*(1), 135–147. doi:10.1016/S0169-409X(01)00245-9 PMID:11755709

Walker, N. J., & Bucher, J. R. (2009). A 21st century paradigm for evaluating the health hazards of nanoscale materials? *Toxicological Sciences*, *110*(2), 251–254. doi:10.1093/toxsci/kfp106 PMID:19468057

Warheit, D. B., Sayes, C. M., Reed, K. L., & Swain, K. A. (2008). Health effects related to nanoparticle exposures: Environmental, health and safety considerations for assessing hazards and risks. *Pharmacology & Therapeutics*, *120*(1), 35–42. doi:10.1016/j.pharmthera.2008.07.001 PMID:18703086

Weaver, S., & Gleeson, M. P. (2008). The importance of the domain of applicability in QSAR modeling. *Journal of Molecular Graphics & Modelling*, *26*(8), 1315–1326. doi:10.1016/j.jmgm.2008.01.002 PMID:18328754

Weissleder, R., Kelly, K., Sun, E. Y., Shtatland, T., & Josephson, L. (2005). Cell-specific targeting of nanoparticles by multivalent attachment of small molecules. *Nature Biotechnology*, *23*(11), 1418–1423. doi:10.1038/nbt1159 PMID:16244656

Westerhoff, P., & Nowack, B. (2013). Searching for global descriptors of engineered nanomaterial fate and transport in the environment. *Accounts of Chemical Research*, *46*(3), 844–853. doi:10.1021/ar300030n PMID:22950943

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Wijnhoven, S. W. P., Peijnenburg, W. J. G. M., Herberts, C. A., Hagens, W. I., Oomen, A. G., & Heugens, E. H. W. et al. (2009). Nano-silver – A review of available data and knowledge gaps in human and environmental risk assessment. *Nanotoxicology*, *3*(2), 109–138. doi:10.1080/17435390902725914

Yuranova, T., Laub, D., & Kiwi, J. (2007). Synthesis activity and characterization of textiles showing self-cleaning activity under daylight irradiation. *Catalysis Today*, *122*(1-2), 109–117. doi:10.1016/j.cattod.2007.01.040

Zahr, A. S., de Villiers, M., & Pishko, M. V. (2005). Encapsulation of drug nanoparticles in self-assembled macromolecular nanoshells. *Langmuir*, *21*(1), 403–410. doi:10.1021/la0478595 PMID:15620331

Zhang, B. T., Zheng, X., Li, H. F., & Lin, J. M. (2013). Application of carbon-based nanomaterials in sample preparation: A review. *Analytica Chimica Acta*, *784*, 1–17. doi:10.1016/j.aca.2013.03.054 PMID:23746402

Zhou, K., Wang, R., Xu, B., & Li, Y. (2006). Synthesis, characterization and catalytic properties of CuO nanocrystals with various shapes. *Nanotechnology*, *17*(15), 3939–3943. doi:10.1088/0957-4484/17/15/055

Zhu, X., Chang, Y., & Chen, Y. (2010). Toxicity and bioaccumulation of TiO_2 nanoparticle aggregates in Daphnia magna. *Chemosphere*, *78*(3), 209–215. doi:10.1016/j.chemosphere.2009.11.013 PMID:19963236

Zhu, X., Tian, S., & Cai, Z. (2012). Toxicity assessment of iron oxide nanoparticles in zebrafish (Danio rerio) early life stages. *PLoS ONE*, *7*(9), e46286. doi:10.1371/journal.pone.0046286 PMID:23029464

Zhu, X., Zhu, L., Chen, Y., & Tian, S. (2009). Acute toxicities of six manufactured nanomaterial suspensions to *Daphnia magna*. *Journal of Nanoparticle Research*, *11*(1), 67–75. doi:10.1007/s11051-008-9426-8

Zhu, X., Zhu, L., Duan, Z., Qi, R., Li, Y., & Lang, Y. (2008). Comparative toxicity of several metal oxide nanoparticle aqueous suspensions to Zebrafish (Danio rerio) early developmental stage. *Journal of Environmental Science and Health, Part A, 43*(3), 278-284.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

KEY TERMS AND DEFINITIONS

%CCT: Consensus percentage or percentage of times that the QSAR-perturbation model correctly predicts the NPs properties. This percentage is used to establish how promising is the QSAR model setup on the virtual screening of different NPs extracted from external sources (*i.e.* usually if it reaches %CCT greater than 50%).

ANN: Artificial neural network machine learning technique.

MAA: Moving average approach used in the QSAR-perturbation based models.

MNPs: Manufactured nanoparticles.

Nano-QSAR Model: Quantitative Structure–Activity Relationships model applied to target the properties of nanomaterials understudy.

Ontology: A set of factors that describe a specific experimental condition under which the properties of NPs have been measured.

Sensitivity of the Descriptors: Relative importance of the variables (descriptors) used in the final QSAR ANN-based model that allows estimating how those should be varied to trigger particular changes in the NPs' properties understudy.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 14
Quantitative Nanostructure–Activity Relationship Models for the Risk Assessment of NanoMaterials

Eleni Vrontaki
NovaMechanics Ltd., Cyprus & University of Athens, Greece

Georgia Melagraki
NovaMechanics Ltd., Cyprus

Thomas Mavromoustakos
University of Athens, Greece

Antreas Afantitis
NovaMechanics Ltd., Cyprus

ABSTRACT

In the last few decades, nanotechnology has been deeply established into human's everyday life with a great number of applications in cosmetics, textiles, electronics, optics, medicine, and many more. Although nanotechnology applications are rapidly increasing, the toxicity of some nanomaterials to living organisms and the environment still remains unknown and needs to be explored. The traditional toxicological evaluation of nanoparticles with the wide range of types, shapes, and sizes often involves expensive and time-consuming procedures. An efficient and cheap alternative is the development and application of predictive computational models using Quantitative Nanostructure-Activity Relationship (QNAR) methods. Towards this goal, researchers are mainly focused on the adverse effects of metal oxides and carbon nanotubes, but to date, QNAR studies are rare mainly because of the limited number of available organized datasets. In this chapter, recent studies for predictive QNAR models for the risk assessment of nanomaterials are reported and the perspectives of computational nanotoxicology that deeply relies on the intense collaboration between experimental and computational scientists are discussed.

DOI: 10.4018/978-1-4666-8136-1.ch014

Copyright © 2015, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

INTRODUCTION

Due to their small size, nanoparticles, exhibit unique physicochemical properties compared to their bulk counterparts and have already found numerous applications in several biomedical and consumer products. Since the production of nanomaterials increases day by day, the probability that they could cause adverse effects to human health and the environment rapidly rises. In order to reduce the cost and time demanded for the experimental evaluation of toxicity, computational methods, such as Quantitative Structure-Activity Relationships (QSARs) are being used to predict the potential toxic effects of nanoparticles. In this chapter, the use of Quantitative Nanostructure-Activity Relationship (QNAR) modeling is discussed as a useful tool to predict the biological and toxicological effects caused by diverse types of manufactured nanoparticles (MNPs). Recent studies are reported summarizing the available predictive QNAR models for the risk assessment of nanomaterials (NMs). Finally, the perspectives and the future of computational nanotoxicology that strongly depend on the intense collaborations between experimental and computational scientists are discussed.

BACKGROUND

According to the U.S. National Nanotechnology Initiative (NNI), nanotechnology is defined as the understanding and control of matter at dimensions of roughly 1 to 100 nanometers, where unique phenomena enable novel applications. At this level, the physical, chemical, and biological properties of materials differ in fundamental and variable ways from the properties of individual atoms and molecules or bulk matter (NNI, 2000).

Nanotechnology has become one of the most promising areas of science and it has fairly emerged as one of the central new technologies in the 21st century, drawing worldwide attention (Haase, Tentschert, & Luch, 2012; Fourches, Pu, & Tropsha, 2011). Hitherto, nanotechnology is an important industrial sector in industrialized countries and it has been referred to as the next industrial revolution. The first national nanotechnology program (NNI) was launched in USA in 2000, to serve as the central point of communication, cooperation, and collaboration for all Federal agencies engaged in nanotechnology research, bringing together the expertise needed to advance this broad and complex field (Bondarenko, 2013). Nanotechnology has found numerous applications in materials science, medicine, cosmetics, or clothing, bioimaging, sensing and electronics (Fourches et al., 2011; Makarucha, Todorova, & Yarovsky, 2011). Nanomedicine, for instance, an emerging application of nanotechnology, provides the possibility of delivering small drug molecules to specific cells, by developing novel diagnostics and therapeutics (Wang, Byrne, Napier, & DeSimone, 2011).

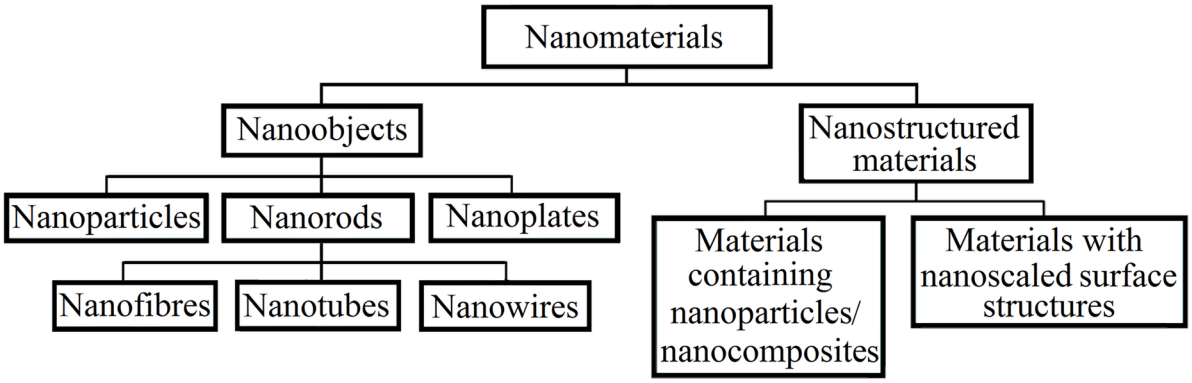
Up to date, several classifications of nanomaterials have been proposed. According to the European Committee for Standardization (CEN) Technical Specification, nanomaterials can be classified into two broad categories, namely nanoobjects and nanostructured materials (Figure 1) (Haase et al., 2012).

According to Hansen, Larsen, Olsen and Baun (2007) and depending on the location of the nanoscale structure in the system, nanomaterials are organized into three broad categories, as it depicted in Figure 2.

For simplicity, nanomaterials can generally be organized into four types (Messina, Olsina, & Stege, 2013):

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 1. Classification of nanomaterials according to CEN/ISO TS 27687; Nanoplates have one external dimension, nanorods two orthogonal dimensions, and nanoparticles all three orthogonal dimensions in the nanoscale.
(ISO 2008).

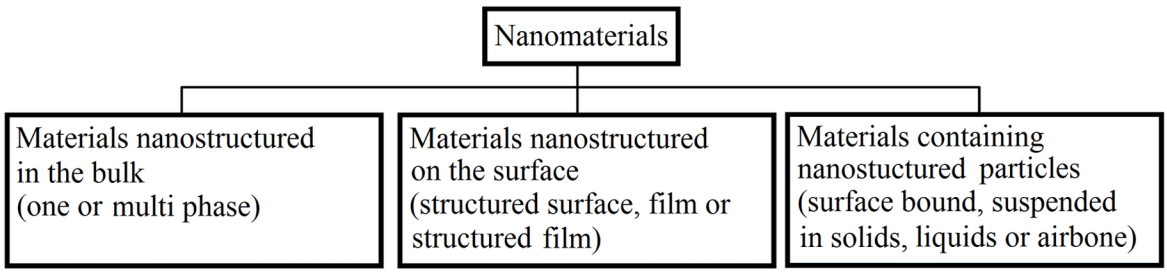


1. Carbon-based materials,
2. Metal-based materials, including metal oxides and quantum dots,
3. Dendrimers, and
4. Composites.

Nanomaterials can be naturally present as dust from forest fires and volcano's activities, biogenic magnetite, viral particles, and even protein molecules, such as ferritin. Nano-sized materials can also be produced unintentionally by anthropogenic sources, such as from combustion by-products. Engineered nanomaterials (ENMs) designed for a specific purpose or function are useful due to their small size and large surface-area to volume ratio, which makes them important as catalysts in chemical reactions, and they have desirable properties as drug delivery devices, as imaging agents in medicine, and in consumer products such as sunscreens, textiles and cosmetics (Clark, White, & Silbergeld, 2011).

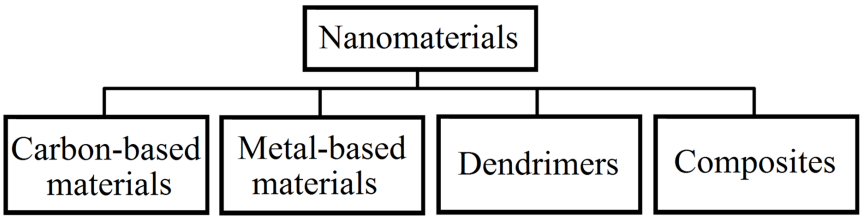
A limited number of nanoparticles are known about their toxicity, including possible mutagenic and/ or carcinogenic effects (Dreher, 2004). Recent studies have shown evident toxicity of some nanoparticles to living organisms, including human body, as well as their potentially negative impact on environmental

Figure 2. Classification of nanomaterials according to Hansen et al.



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 3. Classification of nanomaterials according to Messina et al.



ecosystems and highlight the potential risk related to the development of nanoengineering (Gajewicz et al., 2012). It is expected that nanomaterials can become present in the environment, including air, water and soil, during their life cycle of production, distribution, or usage. In the last decade, a new subdiscipline of ecotoxicology has been developed, that is called nanoecotoxicology with the first scientific papers emerged in the literature in 2006. The ecotoxicological problems derived from the use of NMs and their present in the natural environment and the ecosystems are challenging tasks for environmental toxicologists and ecological risk assessment specialists (Kahru, & Ivask, 2013). It is commonly accepted that there is insufficient knowledge regarding the harmful interactions of nanoparticles with biological systems as well as with the environment.

Since the manufacture and use of nanomaterials are increasing, humans are more likely to be exposed occupationally or via consumer products. Workers in nanotechnology-related industries and small workshops have an increasing risk to be exposed to ENMs. Likewise and as any new technology, the main concerns are about environmental pollution, the toxicity in aquatic organisms and their environmental impact in ecosystems. Lack of available data and low adequacy of experimental protocols prevent comprehensive risk assessment. Toxic effects of nanomaterials on different organisms have already been described (Klaine, 2008; Baun, Hartmann, Grieger, & Kusk, 2008), but measurements of environmental concentrations of ENMs are almost completely absent. The high cost and the long time of the *in vivo* toxicological studies are barriers on testing all of the marketed consumer products incorporating nanoparticles. Our knowledge of the human health effects and environmental concentrations of engineered NMs or NPs is incomplete and also the mechanisms of the interatomic interactions between nanoparticles and biological molecules are not sufficiently understood.

Natural and manufactured/engineered nanoparticles due to their unique properties and biological activity gain the access into the human body through the main ports of body's entry, such as the skin, the gastrointestinal tract and the lungs. Starting from these entries and through the fine capillaries, NPs may also permeate deeper in the organism and tissues, and travel throughout the body interacting with other tissues and organs. Several pharmacokinetic studies have showed that nanoparticles can be found in various cells, in mitochondria, fibroblasts, lipid vesicles, nuclei and macrophages. They have direct contact with cell surface and may lead to the formation of reactive oxygen species (ROS) that results in oxidantive stress and inflammation. Furthermore, they can cause peroxidation of lipids in cellular membranes, resulting in cell damage, and the up-/down-regulation of genes encoding a specific protein involved in inflammatory processes (Gajewicz et al., 2012). Nanoparticles are associated with genotoxicity and neurotoxicity. The effects of nanoparticles on the immune system are still unclear.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Concerning to the ecosystem, there are numerous review articles on the behavior of nanoparticles in aquatic environments (Delay, & Frimmel, 2012), whereas, up to date there are no published reports on the concentration of manufactured nanoparticles in the air (Handy, & Shaw, 2007). Nanoparticles in aquatic systems can be formed by natural abiotic or biotic processes, of natural origin but induced by human activity (e.g. mining), engineered NP that enter aquatic systems incidentally, or engineered NP that are intentionally added to these systems, including nanoscale zero-valent iron ($Fe^{(0)}_{(NP)}$) for removal of As(III) from groundwater (Delay, & Frimmel, 2012). Toxicology studies on animals showed that some types of nanoparticles in the air could cause lung disease.

The potential negative health effects of particles depend on some key parameters, such as the dose, dimension, and durability. Nonetheless, it has been showed that the astonishing physical and chemical properties of nanoparticles that are attributable to their small size differ from these properties of particles at the micro or macro scale (Gajewicz et al., 2012; Nel, Xia, Madler, & Li, 2006). Especially, ultrafine particles are known to have greater adverse health effects than larger particles. In addition, by decreasing for instance the size of gold samples induces color changes from bright yellow through reddish–purple up to blue. The novel and unique properties of nanoparticles (i.e. physicochemical and/or biological characteristics) are often determined by the relationships between their size, surface, and structure. Recent studies have shown that the reduction of particle size increases the surface area and so, smaller nanoparticles have greater surface area (Gajewicz et al., 2012). The toxicity of nanoparticles according to Oberdörster *et al.* (2005) may depend not only on the particle size but on other features with vital importance, such as size distribution, agglomeration state, shape, porosity, surface area, chemical composition, structure-dependent electronic configuration, surface chemistry, surface charge, and crystal structure. To date, the interrelationship between nanoparticles properties and their biological effects remains unclear.

The fact that nanomaterials consist of many substances (primarily composed of carbon, silver, zinc, silica, titanium and gold) and can be found in many sizes and shapes and with a variety of surface coatings, complicates the risk assessment. Studies have showed that nanomaterials after their release of into the environment may tend to agglomeration, aggregation, adhesion, diffusion, dissociation, degradation, adsorption of different species, and bioaccumulation in organisms as well as biomagnifications in trophic pyramids, processes that depend not only on their size/shape, but also on the local environmental and cellular conditions (Gajewicz et al., 2012). Therefore, an assessment of the physical, chemical and biological properties of nanoparticles in each state and case must be made in order to consider their exposure, fate, transport, and transformation in the environment.

Most researchers have focused on the adverse effects to human health and the environment of one single type of particle or several particle types of the same substances, such as,

1. Metal oxide nanoparticles and
2. Carbon nanoparticles, respectively.

Metal oxides (MO_x) is an important group of engineered NPs due to their limited size and a high density of corner or edge surface sites, which allow them wide applications. They can adopt a vast number of structural geometries with an electronic structure that can exhibit metallic, semiconductor or insulator character. They are produced in large quantities and used frequently in cosmetics, sensors and

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

sunscreens, self-cleaning coatings, and textiles. In addition, MO_x's can be used as catalysts for redox reactions in engineered systems and natural environments, water-treatment agents and materials for solar batteries (Puzyn et al., 2011). The most studied oxide systems are Al_2O_3, MgO, ZrO_2, CeO_2, TiO_2 and SiO_2. Less studied nanostructured oxide systems include Zn, Fe, Cu and Sn.

In the last quarter of century, various carbon nanostructures such as fullerenes, carbon nanotubes (CNTs) and graphene have been given great importance in the field of nanoscience and nanotechnology.

Fullerenes are ball shaped hollow structures composed of carbon atoms. The most important fullerene is known C_{60} that has a football-like structure comprising 12 pentagons located around the vertices of an icosahedron and 20 hexagon rings placed at the centers of icosahedral faces (Figure 4). Fullerenes have numerous applications in personal care products, as well as they have several potential uses in medicine.

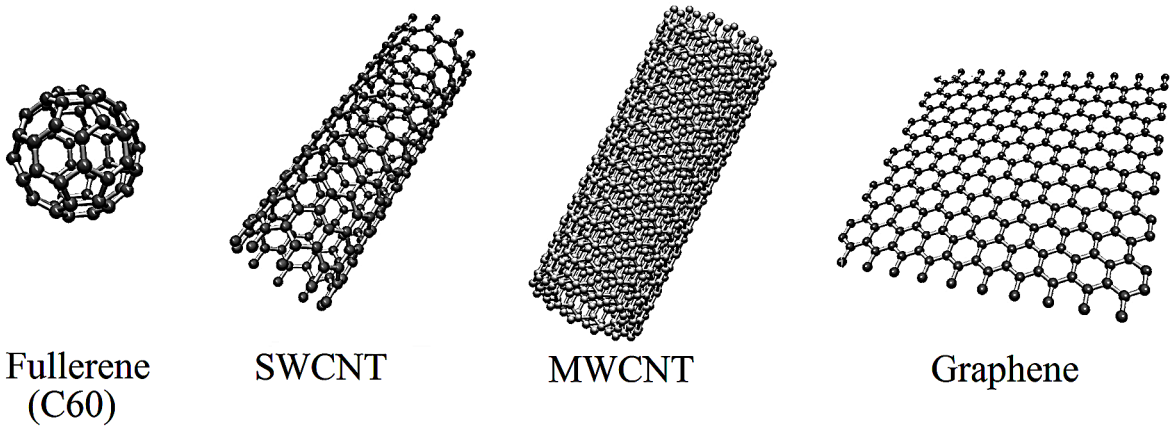
Carbon nanotubes' potential applications in medicine are reported in several studies (Zhou et al., 2008). CNTs can enter various organs and cells, so they can function as intracellular probes, drug carriers, imaging agents, DNA modulators, and other medical devices on the condition that they are biocompatible. There are two structural forms of carbon nanotubes (Figure 4):

1. Single-walled carbon nanotubes (SWCNTs) and
2. Multi-walled carbon nanotubes (MWCNTs).

Graphene, the thinnest but strongest material, is a 2D-sheet composed of carbon atoms characterized by sp^2 hybridization (Figure 4). Some of the applications of carbon nanotubes and graphene are in photodetectors, drug systems, and bioimaging.

It is interesting that the unique properties of metal oxide nanoparticles can be the cause of potential toxic effects. The general results of a study showed that there was a high variation among different metal oxide nanoparticles concerning their ability to cause toxic effects (Karlsson, Cronholm, Gustafsson, & Möller, 2008). It has been showed that nano-TiO_2 causes ROS stress and DNA damage in lymphocytes (Kang, Kim, Lee, & Chung, 2008). Furthermore, similar to TiO_2 nanoparticles, nano-sized silicon oxide (SiO_2), and zinc oxide (ZnO) can induce pulmonary inflammation in rodents and humans (Gordon et al., 1992; Rehn, Seiler, Rehn, Bruch, & Maier, 2003; Chen, Chen, Dong, & Jin, 2004).

*Figure 4. The structures of carbon nanoparticles: fullerene (**C**$_{60}$), single-walled carbon nanotube (SWCNT), multi-walled carbon nanotube (MWCNT), and graphene*



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Similarly, the carbon-based nanomaterials have adverse effects to human's health and ecosystem. Potential short- and long-term toxicity of fullerenes, CNTs and graphene should be carefully investigated before full implementation of these in various biological and commercial applications. Although fullerenes, CNTs and graphene are carbon-based nanomaterials, each of them has different distinctive properties. Different studies have showed that fullerene particles C_{60} suspended in water are characterized by antibacterial activity against *Escherichia coli* and *Bacillus subtilis* (Lyon, Fortner, Sayes, Colvin, & Hughe, 2005) whereas they exhibit cytotoxicity to human cell lines (Sayes et al., 2004). In particular, the cytotoxicity of C_{60} is lower than that of SWCNTs, MWCNTs and graphite (Gajewicz et al., 2012). Therefore, risk assessments of carbon nanomaterials are to be taken on a case-by-case basis to assure high safety.

Quantitative Nanostructure-Activity Relationship (QNAR)

As aforementioned, the traditional toxicological evaluation of the different developmental stages involves expensive, time-consuming and labor-intensive technical procedures. To reduce all above factors, computer-based approaches have been widely developed. The computational methods are a significant tool since they can be utilized to predict the toxicological (biological generally) effects of nanoparticles solely from their physical and chemical properties. Subsequently, by screening *in silico* libraries of virtual nanomaterials, the NMs with the most promising predicted properties can be prioritized. Thus, the costly toxicological studies focus only on a small number of selected and/or rationally designed NPs.

Since 2007, REACH (Registration, Evaluation and Authorization of CHemicals) established by European Parliament and Council, is formally a new chemical management system. One of REACH's key goals is to reduce the number of compounds that require toxicity testing in animals (Schaafsma, Kroese, Tielemans, Van de Sandt, & Van Leeuwen, 2009). REACH legislation has encouraged and foresees the use of alternative *in silico* methods, such as the Quantitative structure-activity relationship (QSAR), to avoid using the animals for the risk assessment of chemicals.

The modeling of manufactured nanoparticles (MNPs) is a major challenge for scientists especially due to the high structural complexity and diversity of MNPs. Therefore, it is very difficult to develop quantitative parameters capable of characterizing the structural and chemical properties of MNPs. The difficulty becomes greater with the absence of experimental data that is required for the validation of computational methods. Puzyn Leszczynska and Leszczynski (2009) first suggested that QSAR modeling can be applied in computational nanotoxicology studies but indicated that no universal QNAR model can accurately assess the toxicity of all possible MNPs.

Similar to general QSAR modeling strategies, the overall objective of QNAR modeling is to develop a mathematical and statistically significant model. To enable the modeling of MNPs, every particle should be described by numerical parameters, called descriptors that are calculated or experimentally measured. The model describes the relationship between the measured biological or toxicological activity profiles of MNPs. The properties such as size, shape, morphology, surface area are usually measured experimentally. Unfortunately, these data for the nanostructures are often absent, so there is an obvious need to develop and employ computational descriptors of MNP structure. To date, thousands of zero-dimensional (0D, referred to atom and bond types counts), 1D, 2D, and 3D (even 4D and 5D!) descriptors have been invented and there are several commercially available software packages (i.e. MOE, DRAGON, CODESSA, SYBYL) to compute several descriptors for any single compound. However and at the moment there are no software packages applicable to MNPs due to their complexity of structure (mixture

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

of organic and inorganic patterns in the same molecule, repetition of structural motifs). Therefore, the descriptors for MNPs have to be measured experimentally until novel and more appropriate types of the descriptors will be developed. To enable QNAR studies, there should be sufficiently large amount of data and a well-defined protocol for model selection and validation should be used.

Models can be built using complex machine learning algorithms, linear or non-linear methods, such as Multiple Linear Regression (MLR), Artificial Neural Networks (ANN), Support Vector Machine (SVM), Random Forest (RF), and k Nearest Neighbors (kNN) or Naïve Bayesian Classifier (NBC), Linear Discriminate Analysis (LDA), and Logistic Regression (LGR).

The QSAR modeling workflow consists of three steps (Figure 5):

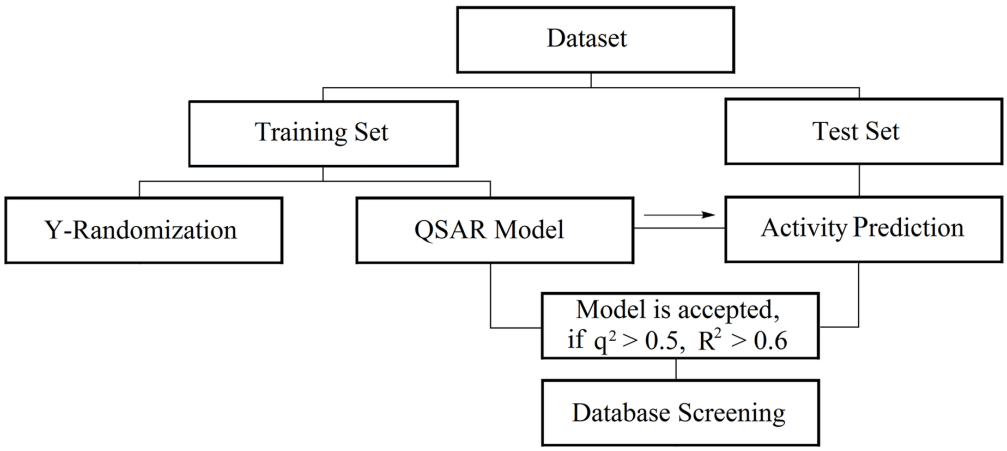
1. Data preparation (compounds and corresponding descriptors),
2. Model building, and
3. Model validation.

Initially, the set of compounds is divided into training and test set. The training set contains the majority of the data, whereas the test set is smaller (~25-30%). The model is derived using only the compounds of training set, and then it is applied to test set to assess their properties. The splitting of the dataset can be done randomly or using a variety of selection techniques. Leave-one-out (or Leave-many-out), cross-validation and Y-randomization are applied to evaluate both prediction and robustness of model. A good model should predict both training and test set (Golbraikh, & Tropsha, 2002) and satisfy some conditions. Finally, the QSAR model can be used increasingly in chemical data mining and combinatorial library design.

A QSAR model to be appropriate validated in a regulatory environment should comply with the OECD guidelines:

1. A clear definition of the endpoint,
2. The use of an unambiguous algorithm,
3. Necessity of defining applicability domain,

Figure 5. The workflow of QSAR modeling



EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

4. Calculating appropriate measures of goodness-of-fit, robustness and predictivity, and
5. Providing a mechanistic interpretation, if possible (retrieved from http://www.oecd.org).

Predicting biological effects of MNPs prior to their synthesis is of outmost importance, and therefore the development of efficient computational approaches is very significant. With the QSAR methodology, if toxicological data are available only for some compounds in a group, the toxicological properties of the other compounds in the same group can also be determined using only calculated molecular descriptors and a suitable mathematical model.

Analyzing data presented in the literature, it is concluded that even if a class of nanoparticles with a similar structure is tested with the same assay and protocol, the number of tested compounds is often insufficient to perform comprehensive internal and external validation of a model and to calculate the appropriate predictivity in QSAR. For instance, Limbach *et al.* (2005) studied the *in vitro* cytotoxicity of seven metal oxide nanoparticles in human and rodent cells and proposed two different rankings: (i) $Fe_2O_3 \approx$ asbestos > ZnO > $CeO_2 \approx ZrO_2 \approx TiO_2 \approx Ca_3(PO4)_2$ and (ii) ZnO > asbestos $\approx ZrO_2 > Ca_3(PO4)_2 \approx Fe_2O_3 \approx CeO_2 \approx TiO_2$, respectively. In another study (Brunner et al., 2006), the same research group found that the formation of reactive oxygen responsible for the toxicity of four metal nanoparticles (TiO_2, Fe_2O_3, Mn_3O_4, and Co_3O_4) toward the human lung epithelial cells was their chemical composition. It must be highlighted that a dataset containing a significant number of compounds is essential to build an appropriately validated QNAR model.

To date, a limited number of QNAR studies and only few reviews concerning the risk assessment of nanoparticles have been reported in the literature (Puzyn et al., 2009; Cattaneo et al., 2010; Burello, & Worth, 2011; Luan et al., 2014). The majority of published QNAR models focus primarily on the predictions of physicochemical properties of nanoparticles (QSPR), e.g. solubility and Young's modulus of carbon-based nanostructures (Toropov, & Leszczynski, 2007; Toropov, Leszczynska, & Leszczynski, 2007; Martin, Maran, Sild, & Karelson, 2007). In this chapter, only the QSAR models for nanotoxicity are reported. In Table 1 eleven datasets available in the literature for QNAR modeling are listed. From these only three satisfy the criterion of relatively large sample size (n>50) and are highlighted in the Table 1.

Table 1. List of available nanoparticles datasets

No.	Dataset	Sample Size	Description	Endpoint
1	Liu et al., 2011	9	Metal oxide NPs	Cytotoxicity
2	Zhang et al., 2012	24	Metal oxide NPs	Cytotoxicity
3	Weissleder et al., 2005	**146**	1 metal core	Cellular uptake in PaCa2
4	Shaw et al., 2008	**51**	4 metal cores	Apoptosis
5	Puzyn et al., 2011	17	Metal oxide NPs	EC_{50} of *E. coli*
6	Sayes et al., 2010	24	MNPs	Cellular membrane damage
7	Gaudin et al., 2012	10	MIL-88B(Fe) MOFs	Caffeine uptake
8	Shinohara et al., 2009	20	fullerene C_{60} NPs	Mutagenicity
9	Zhou et al., 2008	**84**	f-MWCNTs	Protein binding, cytotoxicity, immune responses
10	Tassa et al., 2010	20	MNPs	Ka, Kd of aurora A kinase
11	Luan et al., 2014	41	Metal oxide NPs	Cytotoxicity

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

In 2011, Cohen and his co-workers (Liu et al., 2011) developed a classification-based cytotoxicity nanostructure–activity relationship (nano-SAR) using the high-throughput toxicity screening data for transformed bronchial epithelial (BEAS-2B) cells exposed to nine metal oxide nanoparticles. The model was based on a logistic regression and is validated following accepted QSAR validation guidelines. The High Throughput Screening (HTS) data were first processed to identify and label toxic versus non-toxic events relative to an unexposed control cell population. The best-performing model with 100% classification accuracy was based on three descriptors: the atomization energy of the metal oxide (E_{MeO}), the period of the nanoparticle metal (P_{Me}), and the nanoparticle primary size (d), in addition to the nanoparticle volume fraction (θ_{v}) and was expressed by the following equation:

$$ln\frac{\mathrm{P}\left(NP \in T\right)}{\mathrm{P}\left(NP \in N\right)} = 3600.6 \; + \; 103.5 \; \times \; d \; + \; 9.5 \; \times q_{v} + \; 97.6 \; \times P_{Me} - 58.5 \; \times E_{MeO}$$

where P($NP \in T$) and P($NP \in N$) are the probabilities of a nanoparticle being toxic and nontoxic, respectively.

The authors suggested this successful modeling approach on the basis of the HTS data in order to establish metal oxide nanoparticle hazard ranking. However, they recognized that a significantly larger data set is required in order to improve the reliability of and increase confidence in the nano-SAR approach.

Two years later, nano-SARs for metal oxide NPs toxicity were investigated by the same group using metrics based on dose-response analysis and consensus self-organizing map clustering (Liu et al., 2013). A dataset of 24 metal oxide NPs provided a multiparametric toxicity profile for two different cell lines consisting of seven different toxicity assays (Zhang et al., 2012). Initially, 30 descriptors were used and the best performing nano-SAR model built with support vector machine (SVM) was based on two descriptors. These descriptors, namely conduction band energy (EC) and ionic index (Z^2/r), had a high classification accuracy of ~94%. According to the authors, the class probabilities provided by the nano-SAR enabled the construction of decision boundaries with respect to toxicity classification under different acceptance levels of false negative relative to false positive predictions.

Fourches *et al.* (2010, 2011) developed a QSAR model to predict the cellular uptake of 109 NPs in pancreatic cancer cells (PaCa2). Each NP had the same metal core (iron oxide/NH_2 cores) but diverse surface modifiers; an organic small molecule conjugated to the NP surface (Weissleder, Kelly, Sun, Shtatland, & Josephson, 2005). Accordingly, each NP was represented by a unique set of descriptors determined by the organic molecule. In total, 150 MOE descriptors were calculated. For the developed QNAR model k-nearest neighbors (kNN) method was employed to correlate biological activity of NPs with the calculated descriptors. Their average external prediction power was shown to be high (R^2= 0.72). The authors concluded that QSAR is a useful tool to predict the cellular behavior of a common core-based NP library and their obtained model can be utilized to design a specific coating material to control the cellular uptake.

In the same study (Fourches et al., 2011; Fourches et al., 2010), the researchers studied also a dataset of 51 iron oxide- and quantum dot-based nanoparticles with four different metal cores and surface modifications. The various NPs were tested *in vitro* against four cell lines in different assays by Shaw *et al.* (2008) to study their induced biological effects. The model was built using the support vector machine (SVM)-based classification. An arbitrary threshold was utilized to divide the data set into two groups

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

with the same number of NPs, and four experimental descriptors were used to classify the NPs into two clusters: size, relaxivities, and zeta potential. The model was developed for averaged toxicity responses relative to a control (i.e., unexposed) cell population. The data were classified into two equivalent groups, being below or above an arbitrary threshold. A model classification accuracy of 73% was reported, but without a definitive identification of true toxic end points. The authors suggested that QNAR models can be employed for predicting biological activity profiles of novel nanomaterials, and prioritizing the design and manufacture of a whole NM towards better and safer products.

Similarly, Winklert and his co-workers (Epa et al., 2012) performed a first computational modeling study using the two large sets of experimental nanoparticle data by Weissleder *et al.* (2005) and Shaw *et al.* (2008). They studied all possible combinations of biological assays and cell types, and generated robust models of cellular uptake and induction of apoptosis by metal oxide NPs using Bayesian methods. According to the authors, QNAR is a promising modeling technique, and it can be useful for the analysis of the biological effects of NPs and can capture the effects of different environments of body such as serum, plasma, or lung fluids on nanoparticle composition, as well as NP cellular uptake and interaction with cellular biochemical systems.

Chau and Yap (2012) studied the ability of QNAR models to predict the cellular uptake of Weissleder *et al.*'s NPs dataset with a single metal core in pancreatic cancer (PaCa2) cells. They used four different modeling methods, naïve Bayesian, logistic regression, k-nearest neighbor and support vector machine, to develop candidate models. A final consensus model was developed using the top 5 candidate models, which was able to accurately predict 86.7 to 98.2% of the nanoparticles with good cellular uptake and 67.3 to 76.6% of the NPs with poor cellular uptake. Thus, the authors suggested that it is practical to use the physical, chemical and/or geometrical activity to develop predictive QNAR models.

In 2011, Puzyn *et al.* (2011) investigated the cytotoxicity of 17 metal oxide NPs towards Escherichia coli bacteria. Based on toxicity data they computed structural descriptors. Although the number of NPs used was small, they built a model to predict the cytotoxicity of other nanomaterials. More specifically, the authors calculated 12 descriptors quantitatively describing the variability of the NPs' structure. The descriptors were calculated in MOPAC software package using the quantum-chemical PM6 method. The dataset was divided into training (10 metal oxides) and validation (seven metal oxides) sets. The EC_{50} values of the studied NPs were consequently predicted using the multiple linear regression (MLR). It is notable that Tthe best MLR model was built using only one descriptor selected by genetic algorithm. The model was expressed by the following equation:

$$\log(1/EC_{50}) = 2.59 - 0.50 \times \Delta H_{Me+}$$

where ΔH_{Me+} represents the enthalpy of formation of a gaseous cation in the same oxidation state as that in the metal oxide structure.

The results indicated predictive ability of the model (R^2=0.85, Q^2= 0.77, externally Q^2=0.83). Thus, the authors suggested that their developed QSAR model can be used reliably to predict toxicity of untested nanomaterials, particularly metal oxides in computational nanotoxicology studies.

Leszczynski and his co-workers (Toropov et al., 2012) continued studying the cytotoxicity of the metal oxide NPs to bacteria *Escherichia coli*, and a year later, using CORAL software and the same data set of metal oxide NPs, they boosted the QNAR modeling. In this approach, the pEC_{50} values of the metal oxide nanoparticles' cytotoxicity to bacteria E. coli were used and six random splits into the training and

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

test sets were examined. All the obtained models were satisfactory. According to the authors, CORAL, a convenient application and an available on the Internet software (http://www.insilico.eu/coral), was proved as a reliable tool for the QSAR analysis of the toxicity of metal oxide NPs.

The CORAL software was also used by Toropov *et al.* (2013) in order to build a QSAR model for prediction of cellular uptake of the 109 NPs in pancreatic cancer cells (PaCa2). The software gave satisfactory and stable predictions of the cellular uptake of NPs in PaCa2 cancer cells for five random splits. The optimal molecular descriptors, provided by CORAL, were calculated directly from simplified molecular input-line entry system (SMILES) and are used to correlate with various endpoints. The SMILES gradually has become the most popular representation of the molecular structure in the databases available on the Internet, and SMILES-based optimal descriptors can become attractive alternative to ''classic'' descriptors. Thus, the authors suggested this concept of QSAR as a random event and the alternative to build up QSAR.

Another attempt was made by Ghorbanzadeh, Fatemi, and Karimpour (2012) to build a model for the prediction of the cellular uptake of 109 magnetofluorescent NPs in pancreatic cancer cells. An artificial neural network was employed and the descriptors were calculated using Hyperchem program and Dragon. Six descriptors chosen by combining self-organizing map and stepwise multiple linear regression (MLR) and multilayered perceptron neural network (MLP-NN) modeling techniques were used to correlate the nanostructure of the studied particles with their bioactivity. For the MLR and MLP-NN models, the correlation coefficient and the Root Mean Square (RMS) error were 0.769 and 0.934 and 0.364 and 0.150, respectively. The results obtained after a leave-many-out cross-validation test revealed the credibility of MLP-NN for the prediction of cellular uptake of NPs. Additionally, sensitivity analysis of MLP-NN model indicated that the number of hydrogen-bond donor sites in the organic coating of a NP is the predominant factor responsible for cellular uptake.

Similar to their previous study (Toporov et al., 2013), Toporova and Toporov (2013) worked on TiO_2 nanoparticles. Experimental data on the physicochemical features of TiO_2 nanoparticles and their influence on the membrane damage were taken in the literature (Sayes, & Ivanov, 2010). The optimal descriptors calculated with the Monte Carlo method led to a built up of a predictive model for the membrane damage. The authors suggested that optimal descriptors based on representation of TiO_2 nanoparticles by their physicochemical features can be a tool for modeling of membrane damage.

In a recent study (Sayes, & Ivanov, 2010), nano-SAR classifiers based on nanoparticle properties such as primary size, aggregate size in different media, and zeta potential were developed to predict cell membrane damage induced by TiO_2 and ZnO. The authors compared two approaches to develop mathematical models that have the potential to predict the damage to cellular membranes, as measured by lactate dehydrogenase (LDH) release, of the two metal oxide NPs resulting from several physicochemical features. Given the relatively small number of samples, relatively simple multiple linear regression models were used to probe for possible QSARs related to the endpoint of interest—cellular membrane state. In the cases where the application of the regression models did not produce a good fit, linear discriminant analyses (LDA) classification was applied, and all possible combinations of one, two, and three feature sets were searched and ranked according to how well they are discriminated between different degrees of cellular responses of the biological test system.

A QSAR approach has been undertaken by C. Gaudin *et al.* (2012) for predicting and further interpreting the caffeine uptake, experimentally measured by HPLC and TGA techniques, in a series of functionalized MIL-88B(Fe) materials. Especially, the series of 10 materials consist of MIL-88B(Fe) iron(III) dicarboxylate metal-organic frameworks (MOFs) wherein the organic linker is functionalized by a large variety of polar and non-polar functional groups, such as –H, –Br, –F, $–CF_3$, $–CH_3$, $–NH_2$,

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

–NO_2, and –OH. These MOFs were investigated as potential carriers for encapsulating drugs, using the cosmetic amphiphilic caffeine as a model molecule. The two individual but consensus QSAR models were built using the MLR method and a wide range of molecular descriptors calculated for the series of the investigated grafted organic linkers. The obtained models, validated with the leave-one-out cross-validation and the Y-randomization procedures, led to a good correlation between the observed and calculated values of the caffeine loading for the series of modified MIL-88B(Fe) iron dicarboxylate MOFs.

Recently, a study in computer-aided toxicology of nanomaterials was carried out by Luan *et al.* (2014). The aim of the researchers' work was the prediction of the cytotoxicity of diverse nanoparticles using a QSTR (Quantitative Structure Toxicity Relationship) – perturbation model. The measurements were performed against several mammalian lines and under different experimental conditions, such as the chemical composition of nanoparticles, size, shape and time of exposure of the cell lines. The dataset contained 41 nanoparticles and was retrieved from the literature. Four descriptors were measured;

1. Molar volume (V),
2. Electronegativity (E),
3. Polarizability (P), and
4. Size (L) of nanoparticle.

1681 cases (nanoparticle-nanoparticle pairs) were used and randomly divided into training and prediction sets. The training set included 1261 cases (802 cytotoxic and 459 non-cytotoxic), and the prediction set included the remaining 420 cases (156 cytotoxic and 264 non-cytotoxic). The model was obtained using the LDA method and its statistical quality was examined by Wilks' lambda (λ), chi-square (χ^2), and p-level. The best QSTR-perturbation model was given by the following equation:

$$\mathrm{CTox}_i(c_j)_{\mathrm{nw}} = 1.616\mathrm{CTox}_i(c_j)_{\mathrm{rf}} - 0.229\Delta\Delta V(c_a) - 0.322\Delta\Delta P(c_a) + 0.044\Delta\Delta L(p_s) + 0.502\Delta\Delta V(d_m) - 0.004\Delta\Delta L(t_e) - 0.450$$

According to the authors, the LDA statistics ($\lambda = 0.305$, $\chi^2 = 1490.86$ and $p < 10^{-16}$) showed a very good significance and an accuracy higher than 93% was exhibited for both training and prediction sets. It is of paramount importance that authors in order to prove the significance of the model, predicted correctly the experimentally measured cytotoxicity of new nanoparticles not included in dataset.

Most recently, Kar, Gajewicz, Puzyn, Roy and Leszczynski (2014) using the dataset of Puzyn *et al.*'s (2011) with 17 metal oxides developed and validated a QSTR model. In their effort to encode the cytotoxicity profile of metal oxides to bacteria *E. coli*, they used simple molecular descriptors. More specifically, they calculated seven descriptors based only on the molecular formula and periodic table information:

1. Number of metal atoms (N_{metal}),
2. Number of oxygen atoms (N_{oxygen}),
3. The charge of the metal cation corresponding to a given oxide (χ_{ox}),
4. Metal electronegativity (χ),
5. Sum of metal electronegativity for individual metal oxide ($\sum\chi$),
6. Sum of metal electronegativity for individual metal oxide divided by the number of oxygen atoms present in a particular metal oxide ($\sum\chi/nO$), and
7. Molecular weight (MW).

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

The dataset was randomly divided into training and test sets, including eleven and six metal oxide NPs, respectively. Ten combinations of splitting were performed and the models were developed using stepwise MLR. The best MLR model was built using only one descriptor and expressed by the following equation:

$$pEC_{50} = 4.781 - (1.380 \times \chi_{ox})$$

with $N = 17$, $R^2 = 0.84$, $Q^2 = 0.81$, and $R^2_p = 0.82$.

Despite the obtained statistically significant model, the authors considered it necessary to also perform the partial least squares (PLS) method. More specifically, analyzing the results of MLR they found that for two different metal oxides containing same descriptor value, the predicted value of pEC_{50} was the same. According to authors, the prediction of cytotoxicity of metal oxides using only one parameter had limited ability, and an additional descriptor was needed. The obtained final PLS equation is the following:

$$pEC_{50} = 4.401 - (1.324 \times \chi_{ox}) + (0.176 \times \chi)$$

with $N = 17$, $R^2 = 0.82$, $Q^2 = 0.75$, and $R^2_p = 0.79$.

The authors were satisfied as they succeeded to build up a significant QSTR model using simple descriptors, in no time, and without any software for calculation of descriptors. Furthermore based on this model, Kar *et al.* attempted to provide a most probable mechanism of action of cytotoxicity of metal oxide NPs to bacteria *E. coli*.

Zhou *et al.* (2008) have discovered functionalized multiwalled carbon nanotubes with reduced protein-binding, cytotoxicity, and immune response. They utilized a combinatorial synthesis technique to decorate CNTs' large surface area with a series of organic molecules and investigated the overall toxicity of obtained series of NPs. All CNTs had the same core structure and so their biological/toxicological behavior could depend only on the surface modification. The set of 84 f-MWCNTs was tested in several *in vitro* toxicological assays at two different doses, including protein binding assays, acute toxicity assays, and immune toxicity assays. The authors observed a significant, high stability of experimental measurements for all f-MWCNTs, and suggested further QNAR modeling investigations.

Three years later, Tropsha and his colleagues (Fourches et al., 2011) developed a statistically significant and externally predictive model based on the previous results. Initially, they calculated 2D Dragon descriptors, proceeding to chemoinformatics analysis of the data set. Their findings validated the hypothesis that any f-MWCNT can be modeled using a single, organic molecule from its surface only if all NPs have the same core. They subsequently proceeded to a binary toxicity class to each NP (1 as toxic and 0 as non-toxic) based on the acute toxicity assay and the toxicity measured for CNT with no surface modifiers. 38 strong toxicants (class 1) *versus* 35 weak toxicants (class 0) were subjected to their QNAR random forest (RF) approach and were carried out in a 5-fold external cross-validation. The results were satisfactory and the authors proposed the virtual screening of a library with 240,000 potential CNT surface modifiers in order to identify those materials with desired properties. For inorganic nanoparticles, such as metal oxides, there are some toxicity data available, mainly: EC_{50}, IC_{50}, LC_{50}, LD_{50}, NOEC, zeta-potential, electrophoretic mobility and surface area or solubility. Carbon-based nanoparticles, such as fullerenes, SWCNTs and MWCNTs can also be characterized by available toxicological data, zeta potential and electrophoretic mobility. For carbon nanotubes, there are significantly more data for endpoints, such as surface area, porosity and solubility (Gajewicz et al., 2012). The number of QNAR studies for metal oxide nanoparticles' toxicity is bigger than the one for carbon-based NPs. This may be

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

justified due to the lack of available in the literature datasets of carbon-based nanoparticles with toxicity data. The multiple linear regression (MLR) is the most popular and more applied method of building a QNAR model followed by the support vector machine.

Recently, Toropov and Toropova (2014) worked as in their previous studies (Toporov et al., 2013; Toporova, & Toporov, 2013) also using used optimal descriptors in order to build a QNAR model to predict the mutagenicity of fullerene C_{60} nanoparticles under various conditions using data taken from the literature (Shinohara, Matsumoto, Endoh, Maru, & Nakanishi, 2009). The descriptors were calculated with the Monte Carlo method and the model of 20 fullerene C_{60} NPs was the mathematical function describing the contributions of

1. Dose (g/plate),
2. Metabolic activation, and
3. Illumination (i.e. dark or irradiation).

PERSPECTIVES AND FUTURE OF COMPUTATIONAL NANOTOXICOLOGY

To date, many theoretical studies have been reported on the prediction of mechanical, chemical, and electronic properties of NPs, whereas there have been relatively few published studies on the computational modeling of toxicity. Although quantitative structure–activity relationship (QSAR) methods can play an important role in designing novel and potential drugs and predicting their risk to human health and the environment, the unique properties of nanomaterials and their still unknown toxicity make the implementation of these methods more complicated in comparison with the small organic chemicals. QSAR models usually cover the biological profile of small organic molecules and have been proven accurate in predicting the biological effect for a wide range of molecular scaffolds. QSAR models own a great proportion of their success in the presence of organized databases but no such databases are available for the vast majority of NPs. Experimental data on NPs are scarce and produced by different groups of scientists following different protocols and it is often difficult to select and combine the available information from different sources. On top of that, the structural characteristics of NPs cannot be encoded by the "conventional" widely used 2D and 3D molecular descriptors. NPs include organic as well as inorganic elements with sometimes unknown composition and highly complex structures that demand new approaches for developing molecular descriptors.

In this chapter, we attempted to explain the importance of QNAR modeling and we outlined some of the latest and more significant QNAR studies, especially related to toxicity. As it is demonstrated by several studies QNAR methodologies have emerged as important tools in producing reliable models for the evaluation of MNPs' biological properties and toxicity and can provide valuable information for the design of safer products in nanotechnology.

In an effort to computationally explore data on the effects of nanostructures to human health and the environment, several limitations much be addressed. As already emphasized above nanotechnology, and especially nanotoxicology, is an emerging field of science but unfortunately in this new field of science there is a significant lack of systematic experimental data that characterize structural, geometrical and physicochemical properties of large series of NPs and their effects. To assess risk, regulatory authorities need organized experimental data that are produced to assess properties and effects for a wide range of NPs. When such data are available QNAR studies can then emerge as a useful tool in the identification of the most relevant properties for modeling toxicity of NPs and produce robust and predictive models.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

The availability of quantitative toxicity studies and models is strictly bonded to the availability of high-quality and well organized nanotoxicity data but unfortunately due to the paucity of such data this task is significantly restricted. However, since these limitations have been well recognized, several efforts are now ongoing to select and organize relevant experimental data that are produced under the same protocol and are organized in a common format. For example ISA TAB Nano is a standard tab-delimited format for describing data related to: Investigations, Nanomaterials, Studies (Specimens) and Assays with the goal to develop a specification to facilitate the import/export of data on nanomaterials and their characterizations to/from nanotechnology resources.

It is important to highlight that the conditions and protocols, under which researchers should work, have not been yet clearly defined. There is a variety of *in vitro* measurement techniques and biological endpoints, and each laboratory has its own criteria to test the MNPs. To overcome this barrier, it is required that researchers working in the fields of chemistry, toxicology, biology and chemoinformatics of MNPs collaboratively work together and establish a set of good practices to characterize MNPs and test them in appropriate biological assays under fixed criteria so that data are ready available for in silico exploitation.

As already mentioned, nanomaterials consist of various structures, such as nanotubes, nanorods, fullerenes, crystals, and chemical species of metals, organic compounds, inorganic compounds, etc. This variety could potentially be a problem with the modeling of nanoparticles since we do not know how accurately we can describe the structure of nanomaterials. It is simply impossible to combine all nanoobjects in one description. Since NPs are not structurally homogenous, the properties and the mechanism of toxicity should be studied separately, and thus every group should be analyzed separately. For this reason, sub-classes of structural and physicochemical similarities have been defined and an eight-step procedure is recommended by the Organization for Economic Co-operation and Development (OECD) guidance document on the grouping of chemicals (OECD, 2007).

It is important while developing a QNAR model in order to estimate the potential hazard of ENMs to overcome several common QSAR errors, such as the use of inappropriate tools and methods or use of limited or inappropriate validation procedure. It is important to follow all proposed criteria that can be used to critically assess the suitability of input data and modeling methods and tools for QNAR analysis. Lessons learned from the long standing 'classic' QSAR modeling have to be taken into account from the very beginning when building a QNAR model. It is very important that all five principles recommended by OECD including robust validation and interpretation of results are addressed when the modeling of NPs' biological profile is requested.

Computational methods could address several significant questions emerging from the experimental testing of NPs. An important outcome that needs to be explored is that nanoparticles interact with DNA, proteins, and membranes and establish a series of nanoparticle-biological interfaces. These interactions lead to the formation of protein coronas, particle wrapping, intracellular uptake and biocatalytic processes that could have biocompatible or bioadverse effects. Therefore, it could be a promising approach studying the forces and interactions which occur at the interface between NPs and biological materials to be developed the predictive relationships between structure and activity of nanomaterials.

Up to date, there are several efforts to enable these intensive collaborations between experimentalists, toxicologists, modelers, etc., from both academia and industry. SRC/SEMATECH Engineering Research Program (Retrieved from http://www.erc.arizona.edu/) constitutes a very good example of such collaborative efforts. SRC Engineering Research Center for Environmentally Benign Semiconductor Manufacturing is a dynamic framework with numerous research teams from different U.S. universities and industrial companies. The researchers have synergistic activities by sharing and testing samples of MNPs in different laboratories, creating datasets of them and corresponding QNAR models.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

The NanoPUZZLES project (Retrieved from http://www.nanopuzzles.eu/) is another such of activity which focuses on two groups of compounds:

1. Inorganic engineered nanoparticles (metal oxides NPs), and
2. Carbon-based nanoparticles (carbon nanotubes (single-walled and multi-walled), fullerenes and fullerene derivatives).

The project aims to develop a package of computational algorithms for the comprehensive modeling of the relationships between the structure, properties, molecular interactions and toxicity of selected classes of engineered nanoparticles. The *in silico* approaches that are used to predict the harmful effects of NMs are based on mainly QSARs. In addition, NanoPUZZLES is contributing to the development of an electronic database of well-characterized NMs consisting of physicochemical and toxicological information.

A research project for safe nanotechnology, called NanoMILE and funded by the European Union's Seventh Framework Programme, has been launched from March 2013 (Retrieved from http://www.nanomile.eu-vri.eu/). The aim of this work is the understanding of the mechanisms of nanomaterial interactions with living systems and the environment in order to lead to the development of a single framework for the classification of NM safety and the creation of a universally applicable framework for nanosafety. The MNMs, both metal and carbon-based NMs, will be characterized physicochemically across their entire life cycle and in a wide range of target species. Using a high throughput screening process, a streamlined testing and selection platform will be developed and applied to refine the MNMs selection. The selected MNMs will undergo focused testing relative to their mechanism(s) of effects on living systems and the environment. An iterative experimental/modeling process will integrate the data obtained into quantitative structure or properties/effects relationships.

NANoREG (Retrieved from http://www.nanoreg.eu/) is an enabling project with the cooperation of 60 partners around the Europe. Its aim is to provide the guidance and tools by which a new nano-based material can be tested, evaluated, and approved for commercial use. The partners emphasized that their QSAR modeling should be clearly defined in order to be used as an *in silico* way to predict the behavior of nanaparticles in the human health and environment.

An important aspect that, in the vast majority of examples presented in literature, remains forgotten is the dissemination of the results to the wider community. It is very crucial that the developed model does not remain within the developers' group but is widely disseminated to the community so that it could immediately serve as an important source of information as it was initially designed to be. We should not omit to highlight recent attempts using open source tools for developing *in silico* models for the risk assessment of NMs and moreover the importance of the dissemination of these tools through web applications that could be easily expanded and adjusted to the special needs of each project.

In 2011, Sushko et al. (2011) created a web-based platform that aims to automate and simplify the typical steps required for QSAR modeling. The Online Chemical Modeling Environment (OCHEM) is free for web users and it is available online at http://www.ochem.eu. The OCHEM database was extended to incorporate nanoscale objects and it will be filled continuously with data on the nanotoxicity of metals and metal oxides. Since then, several changes were made to update the OCHEM to be used as a user-friendly database for the nanotoxicity data collection. For the description of the toxic properties of the nanoparticles an abbreviation was used. More specifically, for each property "Nano" prefix was used to separate the nano-properties from all others. The basic characteristics of nanoparticles are chemical composition, average particle size (APS) and shape of the nanoparticles.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Another attempt in this direction is Enalos InSilicoNano platform (Enalos). In this effort KNIME workflows were made available online through Enalos InSilicoNano platform, a web service based solely on open source and freely available software that was developed with the purpose to make the models available to interested user wishing to generate evidence on potential biological effects in the decision making framework. This web service facilitates the computer aided nanoparticles design as it can serve as a source of activity prediction for novel nano-structures. With this tool in hand external databases can be screened within a virtual screening framework to identify novel potent nanoparticles from a prioritized list of compounds. This online tool could be a useful aid for the decision making of both research groups and regulatory bodies interested in NPs' design and screening.

CONCLUSION

It is obvious that nanotechnology has the best prospects to evolve but it should be done in the right way. The unique properties of nanomaterials have several applications in science, and significantly in medicine, but the same properties might result in new pathways and mechanisms of toxic action. For safer nanoscale products QNAR modeling can emerge as a useful tool for the understanding of the relationships between the structure and biological activities of MNPs. To enable safety-by-design using predictive QNAR models, some barriers must be addressed. Datasets on the effects of MNPs need to be enriched and extended with well-defined and consistent experimental measurements. Moreover there is a strong need to supplement the existing set of molecular descriptors with novel "nanodescriptors" that can represent size-dependent properties of nanomaterials. The descriptors have to be both theoretical/computational and experiment-based structural descriptors in order to correlate with biological endpoints. An intense collaboration between the experimentalists (chemists, biologists, toxicologists, etc) and QSAR modelers must be established, and, there is no doubt that the use of QNAR modeling will provide critical support to experimental studies aimed to identify safer nanoparticles with desired properties.

ACKNOWLEDGMENT

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 310451 (Project NanoMILE).

REFERENCES

Baun, A., Hartmann, N. B., Grieger, K., & Kusk, K. O. (2008). Ecotoxicity of engineered nanoparticles to aquatic invertebrates: A brief review and recommendations for future toxicity testing. *Ecotoxicology (London, England)*, *17*(5), 387–395. doi:10.1007/s10646-008-0208-y PMID:18425578

Bondarenko, O., Juganson, K., Ivask, A., Kasemets, K., Mortimer, M., & Kahru, A. (2013). Toxicity of Ag, CuO and ZnO nanoparticles to selected environmentally relevant test organisms and mammalian cells in vitro: A critical review. *Archives of Toxicology*, *87*(7), 1181–1200. doi:10.1007/s00204-013-1079-4 PMID:23728526

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Brunner, T. J., Wick, P., Manser, P., Spohn, P., Grass, R. N., & Limbach, L. K. et al. (2006). In vitro cytotoxicity of oxide nanoparticles: Comparison to asbestos, silica, and the effect of particle solubility. *Environmental Science & Technology*, *40*(14), 4374–4381. doi:10.1021/es052069i PMID:16903273

Burello, E., & Worth, A. P. (2011). QSAR modeling of nanomaterials. *Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology*, *3*(3), 298–306. PMID:21384562

Cattaneo, A. G., Gornati, R., Sabbioni, E., Chiriva-Internati, M., Cobos, E., Jenkins, M. R., & Bernardini, G. (2010). Nanotechnology and human health: Risks and benefits. *Journal of Applied Toxicology*, *30*(8), 730–744. doi:10.1002/jat.1609 PMID:21117037

Chau, Y. T., & Yap, C. W. (2012). Quantitative nanostructure–activity relationship modelling of nanoparticles. *RSC Advances*, *2*(22), 8489–8496. doi:10.1039/c2ra21489j

Chen, Y., Chen, J., Dong, J., & Jin, Y. (2004). Comparing study of the effect of nanosized silicon dioxide and microsized silicon dioxide on fibrogenesis in rats. *Toxicology and Industrial Health*, *20*(1), 21–27. doi:10.1191/0748233704th190oa PMID:15807405

Clark, K. A., White, R. H., & Silbergeld, E. K. (2011). Predictive models for nanotoxicology: Current challenges and future opportunities. *Regulatory Toxicology and Pharmacology*, *59*(3), 361–363. doi:10.1016/j.yrtph.2011.02.002 PMID:21310205

Delay, M., & Frimmel, F. H. (2012). Nanoparticles in aquatic systems. *Analytical and Bioanalytical Chemistry*, *402*(2), 583–592. doi:10.1007/s00216-011-5443-z PMID:22038580

Dreher, K. L. (2004). Health and environmental impact of nanotechnology: Toxicological assessment of manufactured nanoparticles. *Toxicological Sciences*, *77*(1), 3–5. doi:10.1093/toxsci/kfh041 PMID:14756123

Enalos InSilicoNano Platform. (n.d.). Retrieved from www.insilicotox.com

Epa, V. C., Burden, F. R., Tassa, C., Weissleder, R., Shaw, S., & Winkler, D. A. (2012). Modeling biological activities of nanoparticles. *Nano Letters*, *12*(11), 5808–5812. doi:10.1021/nl303144k PMID:23039907

Fourches, D., Pu, D., & Tropsha, A. (2011). Exploring quantitative nanostructure-activity relationships (QNAR) modeling as a tool for predicting biological effects of manufactured nanoparticles. *Combinatorial Chemistry & High Throughput Screening*, *14*(3), 217–225. doi:10.2174/138620711794728743 PMID:21275889

Fourches, D., Pu, D. Q. Y., Tassa, C., Weissleder, R., Shaw, S. Y., Mumper, R. J., & Tropsha, A. (2010). Quantitative nanostructure activity relationship modeling. *ACS Nano*, *4*(10), 5703–5712. doi:10.1021/nn1013484 PMID:20857979

Gajewicz, A., Rasulev, B., Dinadayalane, T. C., Urbaszek, P., Puzyn, T., Leszczynska, D., & Leszczynsk, J. (2012). Advancing risk assessment of engineered nanomaterials: Application of computational approaches. *Advanced Drug Delivery Reviews*, *64*(15), 1663–1693. doi:10.1016/j.addr.2012.05.014 PMID:22664229

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Gaudin, C., Cunha, D., Ivanoff, E., Horcajada, P., Chevé, G., & Yasri, A. et al. (2012). A quantitative structure activity relationship approach to probe the influence of the functionalization on the drug encapsulation of porous metal-organic frameworks. *Microporous and Mesoporous Materials*, *157*, 124–130. doi:10.1016/j.micromeso.2011.06.011

George, S., Xia, T., Rallo, R., Zhao, Y., Ji, Z., & Lin, S. et al. (2011). Use of a high-throughput screening approach coupled with in vivo zebrafish embryo screening to develop hazard ranking for engineered nanomaterials. *ACS Nano*, *5*(3), 1805–1817. doi:10.1021/nn102734s PMID:21323332

Ghorbanzadeh, M., Fatemi, M. H., & Karimpour, M. (2012). Modeling the cellular uptake of magneto-fluorescent nanoparticles in pancreatic cancer cells: A quantitative structure activity relationship study. *Industrial & Engineering Chemistry Research*, *51*(32), 10712–10718. doi:10.1021/ie3006947

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

Gordon, T., Chen, L. C., Fine, J. M., Schlesinger, R. B., Su, W. Y., Kimmel, T. A., & Amdur, M. O. (1992). Pulmonary effects of inhaled zinc oxide in human subjects, guinea pigs, rats, and rabbits. *American Industrial Hygiene Association Journal*, *53*(8), 503–509. doi:10.1080/15298669291360030 PMID:1509990

Haase, A., Tentschert, J., & Luch, A. (2012). Nanomaterials: A challenge for toxicological risk assessment? In A. Luch (Ed.), *Molecular, clinical and environmental toxicology: Environmental toxicology* (pp. 219–250). Heidelberg: Springer. doi:10.1007/978-3-7643-8340-4_8

Handy, R. D., & Shaw, B. J. (2007). Toxic effects of nanoparticles and nanomaterials: Implications for public health, risk assessment and the public perception of nanotechnology. *Health Risk & Society*, *9*(2), 125–144. doi:10.1080/13698570701306807

Hansen, S. F., Larsen, B. H., Olsen, S. I., & Baun, A. (2007). Categorization framework to aid hazard identification of nanomaterials. *Nanotoxicology*, *1*(3), 243–250. doi:10.1080/17435390701727509

Ji, Z., Wang, X., Zhang, H., Lin, S., Meng, H., & Sun, B. et al. (2012). Designed synthesis of CeO_2 nanorods and nanowires for studying toxicological effects of high aspect ratio nanomaterials. *ACS Nano*, *6*(6), 5366–5380. doi:10.1021/nn3012114 PMID:22564147

Kahru, A., & Ivask, A. (2013). Mapping the dawn of nanoecotoxicological research. *Accounts of Chemical Research*, *46*(3), 823–833. doi:10.1021/ar3000212 PMID:23148404

Kang, S. J., Kim, B. M., Lee, Y. J., & Chung, H. W. (2008). Titanium dioxide nanoparticles trigger P53-mediated damage response in peripheral blood lymphocytes. *Environmental and Molecular Mutagenesis*, *49*(5), 399–405. doi:10.1002/em.20399 PMID:18418868

Kar, S., Gajewicz, A., Puzyn, T., Roy, K., & Leszczynski, J. (2014). Periodic table-based descriptors to encode cytotoxicity profile of metal oxide nanoparticles: A mechanistic QSTR approach. *Ecotoxicology and Environmental Safety*, *107*, 162–169. doi:10.1016/j.ecoenv.2014.05.026 PMID:24949897

Karlsson, H. L., Cronholm, P., Gustafsson, J., & Möller, L. (2008). Copper oxide nanoparticles are highly toxic: A comparison between metal oxide nanoparticles and carbon nanotubes. *Chemical Research in Toxicology*, *21*(9), 1726–1732. doi:10.1021/tx800064j PMID:18710264

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Klaine, S. J., Alvarez, P. J. J., Batley, G. E., Fernandes, T. F., Handy, R. D., & Lyon, D. Y. et al. (2008). Nanomaterials in the environment: Behavior, fate, bioavailability, and effects. *Environmental Toxicology and Chemistry*, *27*(9), 1825–1851. doi:10.1897/08-090.1 PMID:19086204

Limbach, L. K., Li, Y., Grass, R. N., Brunner, T. J., Hintermann, M. A., & Muller, M. et al. (2005). Oxide nanoparticle uptake in human lung fibroblasts: Effects of particle size, agglomeration, and diffusion at low concentrations. *Environmental Science & Technology*, *39*(23), 9370–9376. doi:10.1021/es051043o PMID:16382966

Lin, S., Zhao, Y., Ji, Z., Ear, J., Chang, C. H., & Zhang, H. et al. (2013). Zebrafish high-throughput screening to study the impact of dissolvable metal oxide nanoparticles on the hatching enzyme, ZHE1. *Small*, *9*(9–10), 1776–1785. doi:10.1002/smll.201202128 PMID:23180726

Liu, R., Rallo, R., George, S., Ji, Z., Nair, S., Nel, A. E., & Cohen, Y. (2011). Classification NanoSAR development for cytotoxicity of metal oxide nanoparticles. *Small*, *7*(8), 1118–1126. doi:10.1002/smll.201002366 PMID:21456088

Liu, R., Zhang, H. Y., Ji, Z. X., Rallo, R., Xia, T., & Chang, C. H. et al. (2013). Development of structure-activity relationship for metal oxide nanoparticles. *Nanoscale*, *5*(12), 5644–5653. doi:10.1039/c3nr01533e PMID:23689214

Luan, F., Kleandrova, V. V., González-Díaz, H., Ruso, J. M., Melo, A., Speck-Planche, A., & Cordeiro, M. N. D. S. (2014). *Computer-aided nanotoxicology: Assessing cytotoxicity of nanoparticles under diverse experimental conditions by using a novel QSTR-perturbation approach*. Nanoscale. Doi: doi:10.1039/c4nr01285b

Lyon, D. Y., Fortner, J. D., Sayes, C. M., Colvin, V. L., & Hughe, J. B. (2005). Bacterial cell association and antimicrobial activity of a C60 water suspension. *Environmental Toxicology and Chemistry*, *24*(11), 2757–2762. doi:10.1897/04-649R.1 PMID:16398110

Makarucha, A. J., Todorova, N., & Yarovsky, I. (2011). Nanomaterials in biological environment: A review of computer modelling studies. *European Biophysics Journal*, *40*(2), 103–115. doi:10.1007/s00249-010-0651-6 PMID:21153635

Martin, D., Maran, U., Sild, S., & Karelson, M. (2007). QSPR modeling of solubility of polyaromatic hydrocarbons and fullerene in 1-octanol and n-heptane. *The Journal of Physical Chemistry B*, *111*(33), 9853–9857. doi:10.1021/jp071679x PMID:17661505

Messina, G. A., Olsina, R. A., & Stege, P. W. (2013). Integration of nanomaterials in capillary and microchip electrophoresis as a flexible tool. In C. D. García, K. Y. Chumbimuni-Torres, & E. Carrilho (Eds.), *Capillary electrophoresis and microchip capillary electrophoresis, principles, applications, and limitations* (pp. 327–353). Hoboken, NJ: Wiley. doi:10.1002/9781118530009.ch19

NanoMILE. (n.d.). Retrieved from http://www.nanomile.eu-vri.eu/

NanoPuzzles. (n.d.). Retrieved from http://www.nanopuzzles.eu/

NanoReg. (n.d.). Retrieved from http://www.nanoreg.eu/

NanoTox. (n.d.). Retrieved from http://www.nanotox.com/

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

National Nanotechnology Initiative. (n.d.). *Collaborative central point of federal agencies in the USA*. Retrieved from www.nano.gov/about-nni

Nel, A., Xia, T., Madler, L., & Li, N. (2006). Toxic potential of materials at the nanolevel. *Science*, *311*(5761), 622–627. doi:10.1126/science.1114397 PMID:16456071

Oberdörster, G., Maynard, A., Donaldson, K., Castranova, V., Fitzpatrick, J., & Ausman, K. et al. (2005). Principles for characterizing the potential human health effects from exposure to nanomaterials: Elements of a screening strategy. *Particle and Fibre Toxicology*, *6*(2), 8. doi:10.1186/1743-8977-2-8 PMID:16209704

OECD. (2007) *Guidance document on the grouping of chemicals*. Organisation of Economic Cooperation and Development. Retrieved from http://www.oecd.org

Puzyn, T., Leszczynska, D., & Leszczynski, J. (2009). Toward the development of "Nano-QSARs": Advances and challenges. *Small*, *5*(22), 2494–2509. doi:10.1002/smll.200900179 PMID:19787675

Puzyn, T., Rasulev, B., Gajewicz, A., Hu, X. K., Dasari, T. P., & Michalkova, A. 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

Rehn, B., Seiler, F., Rehn, S., Bruch, J., & Maier, M. (2003). Investigations on the inflammatory and genotoxic lung effects of two types of titanium dioxide: Untreated and surface treated. *Toxicology and Applied Pharmacology*, *189*(2), 84–95. doi:10.1016/S0041-008X(03)00092-9 PMID:12781626

Sayes, C., & Ivanov, I. (2010). Comparative study of predictive computational models for nanoparticle-induced cytotoxicity. *Risk Analysis*, *30*(11), 1723–1734. doi:10.1111/j.1539-6924.2010.01438.x PMID:20561263

Sayes, C. M., Fortner, J. D., Guo, W., Lyon, D., Boyd, A. M., & Ausman, K. D. et al. (2004). The differential cytotoxicity of water-soluble fullerenes. *Nano Letters*, *4*(10), 1881–1887. doi:10.1021/nl0489586

Schaafsma, G., Kroese, E. D., Tielemans, E. L. J. P., Van de Sandt, J. J. M., & Van Leeuwen, C. J. (2009). REACH, non-testing approaches and the urgent need for a change in mind set. *Regulatory Toxicology and Pharmacology*, *53*(1), 70–80. doi:10.1016/j.yrtph.2008.11.003 PMID:19061930

Shaw, S. Y., Westly, E. C., Pittet, M. J., Subramanian, A., Schreiber, S. L., & Weissleder, R. (2008). Perturbational profiling of nanomaterial biologic activity. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(21), 7387–7392. doi:10.1073/pnas.0802878105 PMID:18492802

Shinohara, N., Matsumoto, K., Endoh, S., Maru, J., & Nakanishi, J. (2009). In vitro and in vivo genotoxicity tests on fullerene C60 nanoparticles. *Toxicology Letters*, *191*(2-3), 289–296. doi:10.1016/j.toxlet.2009.09.012 PMID:19772904

SRC/SEMATECH Engineering Research Program. (n.d.). Retrieved from http://www.erc.arizona.edu/

Sushko, I., Novotarskyi, S., Körner, R., Pandey, A. K., Rupp, M., & Teetz, W. et al. (2011). Online chemical modeling environment (OCHEM): Web platform for data storage, model development and publishing of chemical information. *Journal of Computer-Aided Molecular Design*, *6*(6), 533–554. doi:10.1007/s10822-011-9440-2 PMID:21660515

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Tassa, C., Duffner, J. L., Lewis, T. A., Weissleder, R., Schreiber, S. L., Koehler, A. N., & Shaw, S. Y. (2010). Binding affinity and kinetic analysis of targeted small molecule-modified nanoparticles. *Bioconjugate Chemistry*, *21*(1), 14–19. doi:10.1021/bc900438a PMID:20028085

Toropov, A. A., Leszczynska, D., & Leszczynski, J. (2007). Predicting water solubility and octanol water partition coefficient for carbon nanotubes based on the chiral vector. *Computational Biology and Chemistry*, *31*(2), 127–128. doi:10.1016/j.compbiolchem.2007.02.002 PMID:17363328

Toropov, A. A., & Leszczynski, J. (2007). A new approach to the characterization of nanomaterials: Predicting young's modulus by correlation weighting of nanomaterials codes. *Chemical Physics Letters*, *433*(1-3), 125–129. doi:10.1016/j.cplett.2006.11.010

Toropov, A. A., & Toropova, A. P. (2014). Optimal descriptor as a translator of eclectic data into endpoint prediction: Mutagenicity of fullerene as a mathematical function of conditions. *Chemosphere*, *104*, 262–264. doi:10.1016/j.chemosphere.2013.10.079 PMID:24246220

Toropov, A. A., Toropova, A. P., Benfenati, E., Gini, G., Puzyn, T., Leszczynska, D., & Leszczynski, J. (2012). Novel application of the CORAL software to model cytotoxicity of metal oxide nanoparticles to bacteria Escherichia coli. *Chemosphere*, *89*(9), 1098–1102. doi:10.1016/j.chemosphere.2012.05.077 PMID:22704203

Toropov, A. A., Toropova, A. P., Puzyn, T., Benfenati, E., Gini, G., Leszczynska, D., & Leszczynski, J. (2013). QSAR as a random event: Modeling of nanoparticles uptake in PaCa2 cancer cells. *Chemosphere*, *92*(1), 31–37. doi:10.1016/j.chemosphere.2013.03.012 PMID:23566368

Toropova, A. P., & Toropov, A. A. (2013). Optimal descriptor as a translator of eclectic information into the prediction of membrane damage by means of various TiO_2 nanoparticles. *Chemosphere*, *93*(10), 2650–2655. doi:10.1016/j.chemosphere.2013.09.089 PMID:24161577

Wang, J., Byrne, J. D., Napier, M. E., & DeSimone, J. M. (2011). More effective nanomedicines through particle design. *Small*, *7*(14), 1919–1931. doi:10.1002/smll.201100442 PMID:21695781

Weissleder, R., Kelly, K., Sun, E. Y., Shtatland, T., & Josephson, L. (2005). Cell-specific targeting of nanoparticles by multivalent attachment of small molecules. *Nature Biotechnology*, *23*(11), 1418–1423. doi:10.1038/nbt1159 PMID:16244656

White, R. M., Sessa, A., Burke, C., Bowman, T., Leblanc, J., & Ceol, C. et al. (2008). Transparent adult zebrafish as a tool for in vivo transplantation analysis. *Cell Stem Cell*, *2*(2), 183–189. doi:10.1016/j.stem.2007.11.002 PMID:18371439

Zhang, H., Ji, Z., Xia, T., Meng, H., Low-Kam, C., & Liu, R. et al. (2012). Use of metal oxide nanoparticle band gap to develop a predictive paradigm for oxidative stress and acute pulmonary inflammation. *ACS Nano*, *6*(5), 4349–4368. doi:10.1021/nn3010087 PMID:22502734

Zhou, H., Mu, Q., Gao, N., Liu, A., Xing, Y., & Gao, S. et al. (2008). A nano-combinatorial library strategy for the discovery of nanotubes with reduced protein-binding, cytotoxicity, and immune response. *Nano Letters*, *8*(3), 859–865. doi:10.1021/nl0730155 PMID:18288815

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

ADDITIONAL READING

Berhanu, W. M., Pillai, G. G., Oliferenko, A. A., & Katritzky, A. R. (2012). Quantitative Structure Activity/Property Relationships: The Ubiquitous Links between Cause and Effect. *ChemPlusChem*, *77*(7), 507–517. doi:10.1002/cplu.201200038

Elzey, S., Larsen, R. G., Howe, C., & Grassian, V. H. (2009). Nanoscience and nanotechnology: Environmental and health impacts. In K. J. Klabunde & R. M. Richards (Eds.), *Nanoscale Materials in Chemistry* (2nd ed., pp. 681–727). Hoboken, NJ: Wiley. doi:10.1002/9780470523674.ch21

Fourches, D., & Tropsha, A. (2014). Quantitative Nanostructure-Activity Relationships: from unstructured data to predictive models for designing nanomaterials with controlled properties. In: N. A. Monteiro-Riviere, & C. Lang Tran (Eds.), Nanotoxicology: Progress toward Nanomedicine, Edition: 2nd. CRC Press (03/2014; ISBN: 9781482203875). doi:10.1201/b16562-11

Gajewicz, A., Puzyn, T., Rasulev, B., Leszczynska, D., & Leszczynski, J. (2011). Metal Oxide Nanoparticles: Size-Dependence of Quantum-Mechanical Properties. *Nanoscience & Nanotechnology-Asia*, *1*, 53–58.

http://www.nature.com/nnano/focus/nanotoxicology

Le, T., Epa, V. C., Burden, F. R., & Winkler, D. A. (2012). Quantitative Structure–Property Relationship Modeling of Diverse Materials Properties. *Chemical Reviews*, *112*(5), 2889–2919. doi:10.1021/cr200066h PMID:22251444

Liu, R., Rallo, R., Weissleder, R., Tassa, C., Shaw, S., & Cohen, Y. (2013). Nano-SAR Development for Bioactivity of Nanoparticles with Considerations of Decision Boundaries. *Small*, *9*(9–10), 1842–1852. doi:10.1002/smll.201201903 PMID:23423856

Marvin, H. J. P., Bouwmeester, H., Bakker, M., Kroese, E. D., van de Meent, D., & Bourgeois, F. (1839). … Verhelst, L. (2013). Exploring the development of a decision support system (DSS) to prioritize engineered nanoparticles for risk assessment. *Journal of Nanoparticle Research*, *15*(8).

Pickrell, J. A., Erickson, L. E., Dhakal, K., & Klabunde, K. J. (2009). Toxicity of inhaled nanomaterials. In K. J. Klabunde & R. M. Richards (Eds.), *Nanoscale Materials in Chemistry* (2nd ed., pp. 730–769). Hoboken, NJ: Wiley. doi:10.1002/9780470523674.ch22

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Rallo, R., France, B., Liu, R., Nair, S., George, S., & Damoiseaux, R. et al. (2011). Self-Organizing Map Analysis of Toxicity-Related Cell Signaling Pathways for Metal and Metal Oxide Nanoparticles. *Environmental Science & Technology*, *45*(4), 1695–1702. doi:10.1021/es103606x PMID:21250674

Sayes, C. M., Smith, P. A., & Ivanov, I. V. (2013). A framework for grouping nanoparticles based on their measurable characteristics. *International Journal of Nanomedicine*, *8*(Suppl. 1), 45–56. doi:10.2147/IJN.S40521 PMID:24098078

Stefaniak, A. B., Hackley, V. A., Roebben, G., Ehara, K., Hankin, S., & Postek, M. T. et al. (2013). Nanoscale reference materials for environmental, health and safety measurements: Needs, gaps and opportunities. *Nanotoxicology*, *7*(8), 1325–1337. doi:10.3109/17435390.2012.739664 PMID:23061887

Thomas, D. G., Gaheen, S., Harper, S. L., Fritts, M., Klaessig, F., & Hahn-Dantona, E. et al. (2013). ISA-TAB-Nano: A Specification for Sharing Nanomaterial Research Data in Spreadsheet-based Format. *BMC Biotechnology*, *13*(1), 2. doi:10.1186/1472-6750-13-2 PMID:23311978

Zhu, Z.-J., Posati, T., Moyano, D. F., Tang, R., Yan, B., Vachet, R. W., & Rotello, V. M. (2012). The Interplay of Monolayer Structure and Serum Protein Interactions on the Cellular Uptake of Gold Nanoparticles. *Small*, *8*(17), 2659–2663. doi:10.1002/smll.201200794 PMID:22730215

KEY TERMS AND DEFINITIONS

Engineered Nanomaterials: Any intentionally manufactured material/particles, designed with specific properties and for specific purpose.

Nanomaterials: The new materials defined as those with one or more dimensions at the nanoscale, which can exhibit novel characteristics compared to the same materials without nanoscale features.

Nanoobjects: Materials with at least one dimension in the nanoscale (1–100 nm).

Nanoparticles: A subset of nanomaterials, with one or more dimensions at the nanoscale level. One of their dimensions has at least length below 100 nm.

Nanoscale: A scale of measurement that uses nanometers (nm) or microns (μm) as units of measure.

Nanostructured Material: Material with an internal or surface structure in the nanoscale.

Risk Assessment: The process of determining the likelihood of a threat or hazard.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chapter 15
QSPR/QSAR Analyses by Means of the CORAL Software:
Results, Challenges, Perspectives

Andrey A. Toropov
IRCCS-Istituto di Ricerche Farmacologiche Mario Negri, Italy

Alla P. Toropova
IRCCS-Istituto di Ricerche Farmacologiche Mario Negri, Italy

Emilio Benfenati
IRCCS-Istituto di Ricerche Farmacologiche Mario Negri, Italy

Orazio Nicolotti
Università degli Studi di Bari "Aldo Moro", Italy

Angelo Carotti
Università degli Studi di Bari "Aldo Moro", Italy

Karel Nesmerak
Charles University in Prague, Czech Republic

Aleksandar M. Veselinović
University of Niš, Serbia

Jovana B. Veselinović
University of Niš, Serbia

Pablo R. Duchowicz
Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas INIFTA (UNLP, CCT La Plata-CONICET), Argentina

Daniel Bacelo
Universidad de Belgrano, Argentina

Eduardo A. Castro
Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas INIFTA (UNLP, CCT La Plata-CONICET), Argentina

Bakhtiyor F. Rasulev
Jackson State University, USA

Danuta Leszczynska
Jackson State University, USA

Jerzy Leszczynski
Jackson State University, USA

DOI: 10.4018/978-1-4666-8136-1.ch015

Copyright © 2015, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

ABSTRACT

In this chapter, the methodology of building up quantitative structure—property/activity relationships (QSPRs/QSARs)—by means of the CORAL software is described. The Monte Carlo method is the basis of this approach. Simplified Molecular Input-Line Entry System (SMILES) is used as the representation of the molecular structure. The conversion of SMILES into the molecular graph is available for QSPR/QSAR analysis using the CORAL software. The model for an endpoint is a mathematical function of the correlation weights for various features of the molecular structure. Hybrid models that are based on features extracted from both SMILES and a graph also can be built up by the CORAL software. The conceptually new ideas collected and revealed through the CORAL software are: (1) any QSPR/QSAR model is a random event; and (2) optimal descriptor can be a translator of eclectic information into an endpoint prediction.

INTRODUCTION

Almost seventy years ago one of the areas of theoretical chemistry named the quantitative structure – property / activity relationships (QSPRs/QSARs) was established (Wiener, 1947, pp. 17-20; Wiener, 1948, pp. 425-430). However, if one considers the publication of Hammett's Free-energy relationship (Hammett, 1935, pp. 125–136) as the first exercise in the QSPR/QSAR analyses, then the introduction of this area of theoretical chemistry should be moved back by eighty years. This is obviously an evidence of the long lasting importance of the QSAR approaches. Such significance is due to their wide applications. The main target of QSPR/QSAR analyses is the prediction of the numerical data related to various endpoints (physicochemical in the case of QSPR; and biochemical, ecological, medicinal, etc., in the case of QSAR).

One can consider three periods of evolution in this field.

1. The first period can be defined as the construction of molecular descriptors and establishing of their correlation with various endpoints (Bonchev & Trinajstic, 1977; Balaban, 1982, pp. 399-404; Bonchev, 1995, pp. 137- 156; Diudea & Gutman, 1998; Ivanciuc, Ivanciuc, & Balaban, 1998; Randic, DeAlba, & Harris, 1998; Castro, Tueros, & Toropov, 2000; Randic & Basak, 2001; Nicolic, Kovacevic, Milicevic, & Trinajstic, 2003; Ren, 2003, pp. 29-39; González, Terán, Teijeira, & González-Moa, 2005; Melagraki, Afantitis, Sarimveis, Igglessi-Markopouloua, & Supuran, 2006).
2. The second period can be defined as the construction of molecular descriptors, establishing of their correlation with an endpoint for the compounds of "visible" training set with the further testing of the predictive potential of this correlation with "invisible" compounds (i.e. compounds which are not involved in building up of the model) involved in the test set (Cronin et al., 2002; Golbraikh & Tropsha, 2002; Taskinen & Yliruusi, 2003; Hemmateenejad, 2004, pp. 475–485; Duchowicz, Castro, Fernandez, & Gonzalez, 2005; Oberg, 2005, pp. 2189–2200; Afantitis et al., 2006; Coi et al., 2006; Leonard & Roy, 2006; Pan, Jiang, & Wang, 2007; Roy & Roy, 2008; Porto, Souza, Junkes, Yunes, & Heinzen, 2008; Toropova, Toropov, Benfenati, Leszczynska, & Leszczynski, 2010; Toropov et al., 2012a; Toropov, Toropova, Raska Jr, Benfenati, & Gini, 2012b; Nesmerak, Toropov, Toropova, & Kohoutova, 2013; Papa, van derWal, Arnot, & Gramatica, 2014).
3. The third period can be defined as building up of the QSPR/QSAR models according to OECD principles (Kruhlak, Contrera, Daniel Benz, & Matthews, 2007; Kar & Roy, 2010, 2012; Putz, Ionascu, Putz, & Ostafe, 2011; de Melo, 2012, pp. 213–222; Sahigara et al., 2012; Toropova & Toropov, 2014).

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

According to the OECD principles, the QSPR/QSAR model should be characterized by:

1. A defined endpoint;
2. An unambiguous algorithm;
3. A defined domain of applicability;
4. Clear measures of predictive potential; and
5. A mechanistic interpretation, if possible (OECD, 2007).

It was during the second period of the methodology development, when optimal descriptors calculated with the molecular graph were suggested as a tool to solve the QSPR/QSAR tasks (Randic & Basak, 1999, 2001; Randic & Pompe, 2001; Toropov & Toropova, 2002; da Silva Junkes, Arruda, Yunes, Porto, & Heinzen, 2005). In fact, majority of molecular descriptors are not expected to correlate with a certain endpoint, but they can be involved in building up model for arbitrary endpoint. However, the optimal descriptor is aimed to correlate with a certain endpoint via a one-variable correlation.

Interestingly, simplified molecular input-line entry system (SMILES) is an attractive alternative for representation of the molecular structure by graph (Weininger, 1988, pp. 31-36; Weininger, 1990, pp. 237-243; Weininger,Weininger, & Weininger, 1989). There are databases available on the Internet where the molecular structure is represented by SMILES (Toropov, Toropova, Raska Jr, Leszczynska, & Leszczynski, 2014). Under such circumstances, the development of QSPR/QSAR approaches, where SMILES is the representation of the molecular structure becomes an attractive direction of research work in the field of the QSPR/QSAR theory and applications. The CORAL software is a tool of the QSPR/QSAR analyses with representation of the molecular structure by SMILES (Toropova, Toropov, Benfenati, Leszczynska, & Leszczynski, 2010d; Toropova et al., 2011a,b). It is to be noted that the CORAL software is able to extract from the SMILES various graph theoretical invariants such as vertex degree and extended connectivity of higher orders (Toropov & Toropova, 2002), as well as invariants for the graph of atomic orbitals (Toropov & Toropova, 2001).

The description of:

1. The organization of the CORAL software,
2. Results obtained using this software,
3. Advantages and disadvantages of the approach, as well as
4. Discussion of possible modifications of this software are the main aims of the chapter.

Background

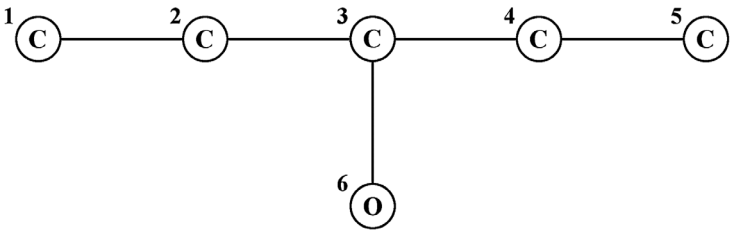
The basic paradigm of the QSPR/QSAR prediction can be expressed as the following:

Endpoint = F(Molecular Structure).

The classic mathematical representation of the molecular structure can be done by means of the adjacency matrix A(G) (Gutman, 1988, pp. 93-94; Gutman, 1997, pp. 281-287). The hydrogen-suppressed graph (HSG) for 3-pentanol (CAS 584-02-1) can be represented as seen in Figure 1.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 1. The hydrogen-suppressed graph for 3-pentanol (CAS 584-02-1)



The hydrogen-suppressed graph is characterized by the following adjacency matrix A(HSG):

$\mathbf{A}\left(\mathbf{HSG}\right)$

	1	2	3	4	5	6	δ_k
1	0	1	0	0	0	0	1
2	1	0	1	0	0	0	2
3	0	1	0	1	0	1	3
4	0	0	1	0	1	0	2
5	0	0	0	1	0	0	1
6	0	0	1	0	0	0	1

$\mathbf{A}^{*}\left(\mathbf{HSG}\right)$

	1	2	**3**	4	5	6	δ_k^*
1	x	1	0	0	0	0	$1+x$
2	1	x	1	0	0	0	$2+x$
3	0	1	x	1	0	1	$3+\boldsymbol{x}$
4	0	0	1	x	1	0	$2+x$
5	0	0	0	1	x	0	$1+x$
6	0	0	1	0	0	y	$1+y$

Majority of the molecular descriptors are a mathematical function of the A(HSG), in particular the Randic index (Randic, 1975) is calculated as

$$^{1}X = \sum_{(i,j)edge}(\delta_i\delta_j)^{-0.5}. \tag{1}$$

The pioneer version of the optimal descriptor (Randic & Basak, 1999; Randic, 2001) is represented by the calculation of classic descriptors using the modified adjacency matrix A*(HSG):

$$^{1}X^{*} = \sum_{(i,j)edge}(\delta_i^*\delta_j^*)^{-0.5}. \tag{2}$$

The next step in the evolution of the optimal descriptors is the application instead of numerical contributions of vertex degrees (δ_k) the so-called correlation weight of the presence of a certain vertex degree, together with the correlation weight of a certain chemical element (A_k):

$$^{0}X_{CW} = \sum_k CW(A_k) + \sum_k CW(\delta_k). \tag{3}$$

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

The correlation weights are calculated by the Monte Carlo method. The numerical data on the correlation weights being used for calculation of the ${}^{0}X_{CW}$ should provide maximum of correlation coefficient between the descriptor and a certain endpoint for the training set of compounds. Having the numerical data one can calculate the ${}^{0}X_{CW}$ for compounds of the training set and define the model:

$$Endpoint = C_0 + C_1 \times {}^{0}X_{CW} \quad (4)$$

However, the predictive potential of the model calculated with Equation 4 should be checked up with external test set.

The representation of the molecular structure in databases by adjacency matrix is not convenient and non economical, in comparison to the representation by SMILES string. Consequently, one can define the descriptor similar to the above-mentioned ${}^{0}X_{CW}$ which can be calculated with SMILES:

$$DCW = F(SMILES) \quad (5)$$

The simplest way is to define the descriptor as a mathematical function of SMILES atoms, i.e. as a mathematical function of each character of SMILES, except situations where two symbols cannot be examined separately (e.g. 'Cl', 'Br', etc.). In the case of 3-pentanol with SMILES= "CCC(O)CC", the scheme can be illustrated as the following:

$$\begin{aligned} &DCW = \\ &CW('C') + CW('C') + CW('C') + CW('(') + CW('O') + CW(')') + CW('C') + CW('C'). \\ &= \Sigma CW(SA_k) \end{aligned} \quad (6)$$

The scheme expressed by Equation 6, can be useful for relatively simple physicochemical properties, such as the normal boiling point of a certain class of organic compounds (alkanes, alkylbenzenes, ethers, etc.). However, in the case of mixture of various classes of organic compounds, or in the case of more complex endpoints (toxicity, carcinogenicity, etc.), the scheme gives poor statistics for both the "visible" training set and for the "invisible" test set. Fortunately, the approach can be improved if in addition to correlation weights of SMILES atoms, combinations of two and / or three SMILES-atoms would be involved in the model.

However, in the cases of two or three SMILES atoms some paradoxical situations occur. For example, SMILES fragment "CN" and "NC" (similarly, "CCN" and "NCC") will be examined as different molecular features. Fortunately, one can avoid these cases if usees normalized combines of SMILES atoms. A simple way is to select order which obey principle: the first and last characters are placed according to their ASCII codes (Toropov, Toropova, & Benfenati, 2010b). In addition, SMILES attributes which are fragments of molecular structure, the global invariants of the molecular structure are involved to build up a model. These global attributes are a mathematical function of presence (absence) of various molecular features which are represented in Table 1.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 1. The scheme of extraction of SMILES atoms and other SMILES attributes in order to build up a model

SMILES Attribute	Example of the Representation for the CORAL Software

S_k	SMILES-atoms, i.e. one symbol or two symbols which cannot be examined separately, e.g. 'C' and 'Cl': this information is represented by sequences of twelve symbols:
	`C..........`
	`Cl.........`

SS_k	A combination of two SMILES-atoms 'CC' and 'CN': this information is represented by sequences of twelve symbols:
	`C...C.......`
	`C...N.......`

SSS_k	A combination of three SMILES-atoms 'CNC' and 'C#N': this information is represented by sequences of twelve symbols:
	`C...N...C...`
	`C...#...N...`

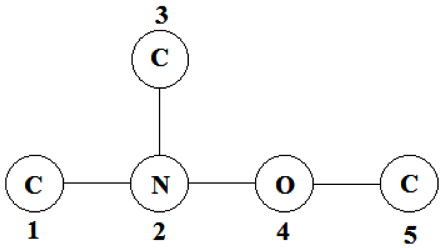
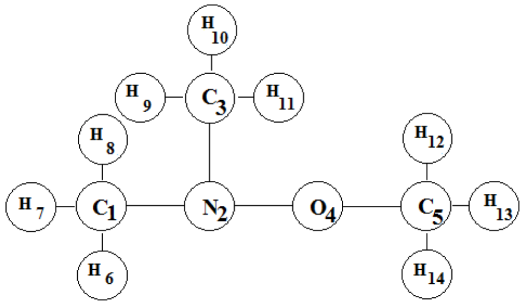
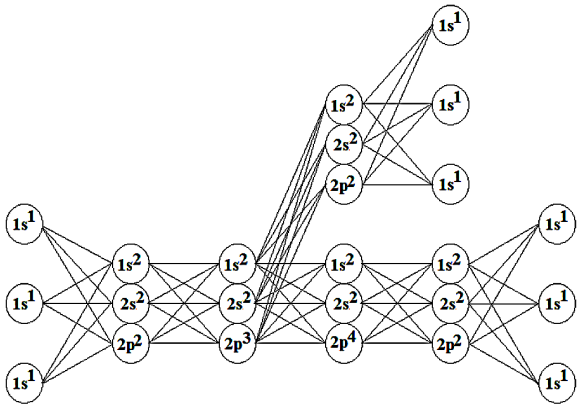
BOND	The presence / absence of double ('='), triple ('#'), and stereo chemical ('@') bonds, e.g. if SMILES = "CCC(O)CC" | = | # | @ | | 0 | 0 | 0 | ⇨ BOND00000000
NOSP	Presence (absence) of nitrogen, oxygen, sulphur, and phosphorus, e.g. if SMILES="CCC(O)CC" | N | O | S | P | | 0 | 1 | 0 | 0 | ⇨ NOSP01000000
HALO	Presence (absence) of fluorine, chlorine, bromine, and iodine atoms, e.g. if SMILES='ClCC(=O)CCl | F | Cl | Br | | 0 | 1 | 0 | ⇨ HALO01000000
PAIR	Simultaneous presence of two SMILES-atoms from the list: F, Cl, Br, I, N, O, S, P, #, =, and @; e.g. if SMILES="ClCC(=O)CCl" the following pairs will be extracted: `++++Cl..O...` i.e. 'Cl' and 'O' `++++Cl..B2..` i.e. 'Cl' and '=' `++++O...B2..` i.e. 'O' and '='

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

The above-mentioned Monte Carlo optimization in the case of the unlimited number of epochs will provide maximum of correlation coefficient for the training set. However (Figure 2), a maximum of correlation coefficient between optimal descriptor and endpoint for a certain number of the epochs exists for the external test set. This amount of epochs is preferable in calculations, because when the number of epochs reaches such a value the obtained model is characterized by a good predictive potential.

Vice versa, the increase of threshold is accompanied by decrease of the correlation coefficient for the training set, but again, the threshold which provide maximum correlation coefficient for the test set exists. This threshold is preferable from practical point of view. Consequently, in order to prepare a good model by means of the CORAL software, one should define preferable values of the threshold and the number of epochs of the Monte Carlo optimization (T* and N*) (Toropova et al., 2011a).

Table 2. The representations of the molecular structure which can be used to build up a model by means of the CORAL software

The Representation of the Molecular Structure	Graphical Interpretation
SMILES	CN(C)OC
Hydrogen-Suppressed Graph, HSG	
Hydrogen-Filled Graph, HFG	
Graph of Atomic Orbitals, GAO	

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 2 contains the list of possible representations of the molecular structure which can be used to build up a model. Thus, the CORAL software is able to provide the following types of QSPR/QSAR models:

1. **SMILES Based Models:**

$$\begin{aligned} &{}^{SMILES}DCW(Threshold, N_{epoch}) = \\ &\alpha\sum CW(S_k) + \beta\sum CW(SS_k) + \gamma\sum CW(SSS_k) + \\ &x \cdot CW(NOSP) + y \cdot CW(HALO) + z \cdot CW(BOND) + t \cdot CW(PAIR) \end{aligned} \quad ; \quad (7)$$

2. **Models Based on the Representation by Molecular Graphs:** These are HSG, or HFG, or GAO (Table 2).

$$\begin{aligned} &{}^{Graph}DCW(Threshold, N_{epoch}) = \sum CW(A_k) + \\ &\alpha\sum CW({}^0EC_k) + \beta\sum CW({}^1EC_k) + \gamma\sum CW({}^2EC_k) + \delta\sum CW({}^3EC_k) \end{aligned} \quad ; \quad (8)$$

In Equations 7 and 8, coefficients α, β, γ, δ, t, x,y, and z can be 1 (yes) or 0 (no). Various combines of their values give possibility to define certain version of the optimal descriptor;

3. **The Hybrid Models which Involves Molecular Feature Extracted from Both Smiles and Graph:**

$${}^{Hybrid}DCW(Threshold, N_{epoch}) = {}^{SMILES}DCW(Threshold, N_{epoch}) + {}^{Graph}DCW(Threshold, N_{epoch}) \quad ; \quad (9)$$

4. Last but not least: the CORAL software is able to provide models for phenomena where eclectic information is the only basis for prediction, e.g. property and activity of proteins (Toropov, Toropova, Raska Jr, Benfenati, & Gini, 2012b) as well as endpoint related to nanomaterials (Toropova & Toropov, 2013). The generalized form of the optimal descriptor which can be used for the above-mentioned situations is the following:

$$DCW(Threshold, N_{Epoch}) = \Sigma CW(CI_k) \quad (10)$$

where CI_k represent codes of available eclectic data (e.g. concentration, size, condition of synthesis, technological details, etc.).

Selected Technical Details Related to the CORAL Software

There are some limitations of the SMILES in CORAL software. The maximal length for SMILES is 500 symbols; the maximal number of atoms (or atomic orbitals in the case of the graph of atomic orbitals) is equal to 300; the maximal number of compounds (which are involved in building up model) is equal to 50,000.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Development of a Model by Means of the CORAL Software Is Based on the Following Principles

1. Molecular structure of each compound can be represented by molecular features which are extracted from SMILES and/or from molecular graph;
2. There are local and global molecular features which can be extracted from SMILES. The local features represent some fragments (or individual atoms). The global features are some indexes which characterized molecules as whole.
3. Building up of QSPR/QSAR model for an arbitrary split into the training and validation sets can be examined as a random event.
4. The statistical quality of each QSPR/QSAR model is a mathematical function that represents split of the data into the "visible" training set and "invisible" test set.
5. The average statistical quality of QSPR/QSAR models that is obtained for several splits into training and test sets is more robust criterion for the estimation of an approach than the statistical quality for solely one split.
6. The average statistical quality of a model for external "invisible" validation sets is more significant information related to the quality of the model than the average statistical quality for "visible" (i.e. substances involved in building up model) sub-training and calibration sets.

Every model that includes geometry-dependent molecular descriptors usually involves a relatively expensive calculation of the optimum molecular geometry, involving high computational costs and long CPU time. In this context, the conformation-independent 0D, 1D and 2D-QSPR methods emerge as an alternative approach for developing models based on constitutional and topological molecular features of compounds (Duchowich, Comelli, Ortiz, & Castro, 2012; Talevi et al., 2012). The exclusion of 3D-structural aspects also avoids problems associated with ambiguities, resulting from an incorrect geometry optimization output due to the existence of compounds in various conformational states. Such kind of problems may also lead to the vanishing of predictive capability of the QSPR/QSAR when applied for the prediction of an external test set of compounds. In the realms of the CORAL approach, the calculated flexible descriptor is a molecular descriptor which depends both on the molecular structure and the property under analysis, but does not explicitly depend on details from the 3D-molecular geometry. In previous QSPR/QSAR studies the importance of this methodology has been shown. Such approach is able to provide models having a comparable or sometimes better quality than the ones found by searching the best descriptors in a pool containing thousands of 0D-3D descriptors (Mullen, Duchowicz, & Castro, 2011**;** Garcia et al., 2011; Ibezim, Duchowicz, Ortiz, & Castro, 2012).

THE COLLECTION OF QSPR/QSAR MODELS

Over the years the CORAL software has been used for QSPR/QSAR analyses of a group of various endpoints. The results are listed in the following sections.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Water Solubility of Organic Compounds

Data on water solubility of 1311 substances, i.e. their CAS number, SMILES, and the values of negative logarithm of water solubility –log S (mol / L) were taken from the web site of Virtual Computational Chemistry Laboratory (http://www.vcclab.org/lab/alogps/). These substances were distributed by means of five random splits into the sub-training set (≈55%), calibration set (≈25%), and test set (≈20%). The statistical quality of the CORAL models (Table 3) is better than the statistical quality of the models suggested earlier in the literature for the same large database (Huuskonen, 2000, pp. 773-777; Tetko, Tanchuk, Kasheva, & Villa, 2001; Yan & Gasteiger, 2003). The best prediction is characterized by n_{test}=21, r^2_{test}=0.91, s_{test}=0.63 (Huuskonen, 2000, pp. 773-777).

Solubility of [C60] and [C70] Fullerene Derivatives

The numerical data on the solubility of [C60] and [C70] fullerene derivatives in chlorobenzene in mg/mL were taken from the literature (Troshin et al., 2009). Unfortunately, since at present, the QSPR analysis of these data by other methods is not available we are not able to compare the obtained results with various approaches.

Octanol/Water Partition Coefficient of Antineoplastic Drugs

The numerical data on octanol/water partition coefficient (logP) for 55 organic compounds have been taken from the literature (Wishart at al., 2008). The CORAL model (Table 3) is better than the model described in the literature (Fouchécourt, Béliveau, & Krishnan, 2001) for the same data (n=55, r^2=0.8180).

Half-Wave Potential

The half-wave oxidation/reduction potential is very useful characteristic of electronic properties of the chemical compounds. The first SMILES-based QSPR analysis of half-wave potentials was performed for 40 benzoxazines (Toropov, Nesmerak, Raska Jr., Waisser, & Palat, 2006). The statistical quality this approach (r^2 = 0.882) is fully comparable with classical approach based on Hammett constants (r^2 = 0.897; Nesmerak, Nemec, Sticha, Waisser, & Palat, 2005). Other example of SMILES-based QSPR of half-wave potential using CORAL software is analysis of experimental data for 23 derivatives of *N*-benzylsalicylthioamides. From the statistical point of view, the model is comparable with data obtained in the experiment (Nesmerak, Toropov, Toropova, Kohoutova, & Waisser, 2013).

Rate Constants of Reactions between Organic Aromatic Pollutants and Hydroxyl Radical

The degradability of 78 compounds expressed in terms of second-order degradation rate constants, K_{OH}, between $OH^{\cdot}$- and each targeted aromatic compound have been taken from the literature (Kušić, Rasulev, Leszczynska, Leszczynski, & Koprivanac, 2009). Decimal logarithm of the K_{OH} was evaluated as the endpoint for the QSPR models. The model from the above-mentioned work (Kušić, Rasulev, Leszczynska, Leszczynski, & Koprivanac, 2009) is characterized by r^2_{test}=0.760. Hence the statistics of the CORAL model (Table 3) are better, because for four random splits into the training and test sets for the CORAL models (Toropov et al., 2012a) the average r^2_{test}=0.8253.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Table 3. Statistical quality of QSPR/QSAR models for various types of property/activity for selected sets of data calculated with the CORAL software

Endpoint	N_{train}	r^2_{train}	s_{train}	N_{test}	r^2_{test}	s_{test}	R_m^2	References
Physicochemical Parameters								
Water solubility of organic compounds	721	0.9231	0.565	261	0.9381	0.511	0.8963	Toropov et al., 2013a
Solubility of [C60] and [C70] fullerene derivatives	18	0.758	17.6	9	0.925	12.5	0.902	Toropova et al., 2011b
Octanol/water partition coefficient of antineoplastic drugs	22	0.9306	0.601	12	0.9776	0.727	-	Toropov et al., 2010a
Half-wave potential	9	0.7333	0.021	5	0.7295	0.028	0.8294	Nesmerak et al., 2013
Rate Constants of Reactions Between Organic Aromatic Pollutants and Hydroxyl Radical	38	0.8207	0.183	20	0.8033	0.145	0.6731	Toropov et al., 2012a
Biochemical Activities								
Toxicity towards *Daphnia magna*	149	0.7006	1.04	89	0.7680	0.878	0.7413	Toropova et al., 2012a
Toxicity in rats	344	0.7662	0.468	126	0.7237	0.315	0.6388	Toropova et al., 2011a
Toxicity of binary mixtures	14	0.9584	0.167	10	0.9362	0.200	0.7164	CORAL: Models of toxicity of binary mixtures. Chemometrics and Intelligent Laboratory Systems, 119
Carcinogenicity	170	0.6278	0.871	61	0.7576	0.602	-	Toropova et al., 2010b
Carcinogenicity of drugs	1088	0.8700	1.46	376	0.8520	1.53	0.835	Toropova & Toropov, 2014
Anticancer agents Murine P388 Leukemia	75	0.7688	0.48	25	0.8025	0.49	-	Toropov et al., 2010c
Anticancer agents SK-MEL-5 cell line UACC-62 cell line	19 19	0.9506 0.9409	0.14 0.13	6 6	0.752 0.760	0.36 0.40	-	Mullen et al., 2011
Anti-malaria agents aryl-piperazine derivatives	16	0.9960	0.040	6	0.8836	-	-	Ibezim et al., 2012
Anti-malaria agents 4(1-H)-quinolone derivatives	20	0.7009	0.575	10	0.6130	0.89	0.55	Toropov et al., 2013b
Anti-HIV-1	7	0.8413	0.309	5	0.8379	0.642	-	Toropov et al., 2010d
High-affinity 5-HT1A receptors	40	0.9457	0.223	18	0.9473	0.298	0.8325	Veselinovic et al., 2013a
Calcium Channel-Antagonistic Effect	50	0.8507	0.706	22	0.9269	0.521	0.8558	Veselinovic et al., 2013b
Anticonvulsant agents	46	0.7627	0.192	5	0.7001	0.313	0.6428	Garro Martinez et al., 2011
Alkaloid toxicty	26	0.83	0.19	8	0.70	0.44	-	Turabekova et al., 2014
Membrane damage TiO_2 nanoparticles	10	0.9893	0.025	5	0.9647	0.066	0.90	Toropova & Toropov, 2013

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chromatographic Retention Data

The retention factors of 41 derivatives of 1-phenyl-5-benzylsulfanyltetrazole in reverse-phase HPLC were measured for various compositions of mobile-phase. The retention characteristics of the linear solvent strength model were calculated using the linear relationship between logarithm of retention factor and volume fraction of organic modifier in mobile phase. The logarithms of the solute retention factor extrapolated to a mobile phase composition with 0% organic modifier, and constants for a given solute in a given chromatographic system (i.e. the slopes of the linear regressions) were used as endpoints. The one-variable QSPRs developed show that the CORAL software is a tool for building up a robust quantitative structure-retention relationships (QSRR) model for retention characteristics (Nesmerak, Toropov, & Toropova, 2014).

Toxicity towards *Daphnia magna*

The descriptions of toxicity of organic chemicals related to 48 h *Daphnia magna* toxicity expressed in negative decimal logarithm of the dose that kills 50% of organisms i.e. pLC50 were taken from the literature (Kar & Roy, 2010). The data set covers range of octanol/water partition coefficient from −2 to 8. According the values, it could be more possibly logarithms of octanol/water partition coefficient. The range of toxicity (daphnia) is from 0.46 to 10.09. In regard to the chemical domain, the data set includes hydrocarbons, aliphatic alcohols, phenols, ethers, and esters; anilines, amines, nitriles, nitroaromatics, amides, and carbamates; urea and thiourea derivatives; iso-thiocyanates; thiols; phosphorothionate and phosphate esters; and halogenated derivatives. The statistical quality of the model described in the original article is the following: n=222, r^2=0.695 (training set), and n=75, r^2_{pred}=0.741, R_m^2 =0.707 (test set). Thus, the statistical quality of the Coral model (Table 3) is at least comparable with the model suggested in the above-mentioned work (Kar & Roy, 2010).

Toxicity in Rats

Rat toxicity data (LD50, in mg/kg, oral exposure, n=689) was taken from the U.S. Library of Medicine (http://toxnet.nlm.nih.gov/). The $\log_{10}$[1/LD50] has been used as endpoint. Unfortunately, information on QSAR analysis of these data by other approaches is not available. However, comparison of optimal descriptors and Multiple Linear Regression (MLR) has shown (Toropov, Rasulev, & Leszczynski, 2008) that the models for rat toxicity based on optimal descriptors calculated with SMILES have better statistical characteristics than one-, two-, three-, and four-variable MLR models based on the topological and quantum mechanical descriptors.

Toxicity of Binary Mixtures

The numerical data on the toxicity of binary mixtures was taken from the literature (Zhang, Zhou, Yang, & Wang, 2007). The toxicity is expressed as pEC50 (i.e. negative decimal logarithm log[1/EC50]), logarithm of the inverse of the effective concentration required to bring about a 50% decrease in light emission, for *Photobacterium phosphoreum* (T3 mutation). The statistical quality of four-variables model (calculated with involvement of the quantum mechanical descriptors) suggested in the literature (Zhang, Zhou, Yang, & Wang, 2007) for the toxicity of the same 50 binary mixtures is the following: n=50, r^2=0.85, s=0.270. Consequently, the CORAL model for this endpoint provides better results (Table 3).

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Carcinogenicity

Experimental values for carcinogenicity were taken from the web site of United States Environmental Protection Agency (http://www.epa.gov/ncct/dsstox/sdf_cpdbas.html). Carcinogenicity is expressed as the potency dose that induces cancer in rats (TD50, in mg/kg body weight). These values have been converted into mmol/kg body weight. The log(TD50) was selected as endpoint for modeling. The correlation coefficient of the QSAR for carcinogenicity of 45 benzene derivatives amounts to about r^2 = 0.7 (Vracko, 1997, pp.1037–1043). Another QSAR model for rodent carcinogenicity nitroso compounds obtained by multiple linear regression analysis (Helguera, Cordeiro, Pérez, Combes, & González, 2008) has the following statistical characteristics: n=48, r^2 = 0.859, s=0.361, F=42 (training set) and n=6, Q^2 = 0.71, s=0.488 (test set). The CORAL model for carcinogencity (Table 3) should be consider as comparable with the above-mentioned models, since the number of examined compounds for the CORAL model is considerable larger.

Carcinogenicity of Drugs

Numerical data on carcinogenic potentials of the selected 1464 organic compounds (chemical domain which includes hydrocarbons, aliphatic alcohols, phenols, ethers, and esters; anilines, amines, nitriles, nitroaromatics, amides, and carbamates; urea and thiourea derivatives, isothiocyanates, thiols, phosphate esters, and halogenated derivatives) are expressed by DF (Discriminate Function). The range of DF is from -9.91 to 9.86. Positive value of DF is an indicator of carcinogenic compound, negative value of DF is an indicator of non-carcinogenic compound. The predictive potentials of QSAR models for the same series of compounds suggested in the literature are characterized by n_{test}= 732, r^2_{test}=0.713 (Kar & Roy, 2011); and n_{test}= 732, r^2_{test}=0.77 (Duchowicz, Comelli, Ortiz, & Castro, 2012). Thus the statistical quality of the CORAL model is better than those available in the literature (Table 3).

Anticancer Agents (Murine P388 Leukemia)

The decimal logarithm log(1/IC50), where IC50 represents the concentration of the agent necessary to reduce cell viability by 50% against Murine P388 Leukemia (in vitro cytotoxic activity) is considered as the endpoint. Numerical data on this endpoint was taken from the literature (Atanasova, Ilieva, & Galabov, 2007). The QSAR model suggested in the published work (Atanasova et al., 2007) is characterized by n_{test} = 25; r^2_{test}=0.700. Thus, predictive potential of the CORAL model for the anticancer activity is better that the previously published results.

Anticancer Agents (SK-MEL-5 Cell Line and UACC-62 Cell Line)

The observed cell anti-proliferative activities of the triphenylmethyl containing compounds inducing death in the human melanoma cell lines SK-MEL-5 and UACC-62 are extracted from the literature (Palchaudhuri, Nesterenko, & Hergenrother, 2008; Palchaudhuri & Hergenrother, 2008). The models calculated with MLR approch (Mullen, Duchowicz, & Castro, 2011) are characterized by n_{train}=17, r^2_{train}=0.781, s_{train}=0.16, n_{test}=6, r^2_{test}=0.243, s_{test}=0.44 (SK-MEL-5); and n_{train}=19, r^2_{train}=0.859, s_{train}=0.17, n_{test}=4, r^2_{test}=0.722, s_{test}=0.30 (UACC-62). It can be compared to the results obtained using CORAL approach, included in the Table 3.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Anti-Malaria Agents (Aryl-Piperazine Derivatives)

The experimental inhibitory concentrations (IC50) in micromolar units of aryl-piperazine derivatives against the chloquine resistant strains W2 and FCR3 and against the chloroquine-sensitive strains D10 and NF54 are extracted from the literature (Molyneaux, Krugliak, Ginsburg, & Chibale, 2005). The QSAR calculated with optimal descriptors gave possibility to suggest some perspective molecular structures of anti-malaria agents.

Anti-Malaria Agents (4(1-H)-Quinolone Derivatives)

A series of 53 endochin analogs (4(1-H)-quinolone derivatives) with anti-malarial activity against the clinically relevant multidrug resistant malarial strain TM-90-C2B has been taken from the literature (Ojha & Roy, 2011). The statistical characteristics of the model suggested in the published work (Ojha & Roy, 2011) are the following: $n_{train} = 39$, $r^2_{train} = 0.797$, $s_{train} = 0.517$; $n_{test} = 14$, $r^2_{test} = 0.808$. Thus, the statistical quality of the CORAL model is comparable with the above-mentioned model (Ojha & Roy, 2011).

Anti-HIV-1

The binding affinity values of fullerene derivatives pEC50 were taken from the literature (Durdagi, Mavromoustakos, & Papadopoulos, 2008). The QSAR analysis of these data has shown that even for the case of small set of data, the Monte Carlo technique gives reproducible results. In spite of relative lower statistical quality of the CORAL model for the endpoint, this model provides satisfactory prediction for a several splits of the data into the training and test sets. Thus, the optimal descriptors should be considered as reasonable alternative for the 3D QSAR analysis.

High-Affinity 5-HT1A Receptors

A dataset of 88 arylpiperazines considered as high affinity 5-HT1A receptor ligands, to which the in vitro affinity values (as measured by inhibition constants, Ki) were collected from the literature (Martinez-Esparza et al., 2001a,b) was analyzed. The CORAL software is able to be an efficient tool to build up a robust model for affinity of 5-HT1A receptor ligands of arylpiperazines. The suggested modeling process for affinity of 5-HT1A receptor ligands is based on the representation of the molecular structure by SMILES and on the application of experimental data. The predictive potential of the applied approach was tested with four random splits of the data into the training and test sets. In addition, the SMILES attributes, which are promoters of increase/decrease ligand binding affinity to 5- HT1A receptor, were identified. The mechanistic interpretation of these models from probabilistic point of view was suggested.

Calcium Channel-Antagonistic Effect

A dataset of 72 1,4-dihydropyridine derivatives with values for pIC50 was taken from the literature (Gecen, Saripinar, Yanmaz, & Şahin, 2012). Almost identical statistical quality of the QSAR models has been obtained for four different random splits of data. The mechanistic interpretation of these models in terms of promoter increase / decrease of the endpoint was suggested.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Anticonvulsant Agents

The experimental information on the antiepileptic activities ED50 of the molecular structures was obtained from the literature (Eddington et al., 2002, 2003; Edafiogho, Ananthalakshmi, & Kombian, 2006; Cox, Gao, Raje, Scott, & Eddington, 2001a; Cox, Scott, Gao, Raje, & Eddington, 2001b; Vamecq, Lambert, Poupaert, Masereel, & Stables, 1998). The MLR analysis gives preferable accuracy of the two ED50 models (Garro Martinez, Duchowicz, Estrada, Zamarbide, & Castro, 2011) r^2_{test}=0.856 and r^2_{test}=0.948. For this case the CORAL model provides worse results (Table 3).

Anesthetic Activity of Alkaloids

The duration of anesthesia (related to protein binding of a drug) and the onset time (determined by the pK_a) are important characteristics in assessment of local anesthetic agents. In the study (Turabekova et al., 2014) we utilized toxicity, duration, and onset of action as endpoints to construct QSAR models for the series of 34 diterpenoid alkaloids characterized by local anesthetic activity using GA-MLRA/PLS and SMILES-based optimal descriptors approach (CORAL). SMILES-based optimal descriptor approach application resulted in models of relatively improved both statistical fit of toxicity and duration of anesthesia models (r^2≈85% and r^2≈95%, respectively) and predicting ability (≈72%) for all three endpoints. This approach appeared to be more sensitive to structural peculiarities of molecules than regression methods that have to be taken into account in potential data splitting.

Membrane Damage TiO_2 Nanoparticles

Large databases related to physicochemical and biological activity of nanomaterials remain scarce. Under such circumstances, the possible way to build up a predictive model is an analysis of eclectic available data on various phenomena related to nanomaterials. As a result of exchange of classic paradigm Endpoint = F (molecular structure) by paradigm Endpoint = F(eclectic information), an analogy of classic QSPR/QSAR approach applied to nanomaterials can be obtained (Toropova & Toropov, 2013). The data obtained using this approach are provided in the Table 3.

SOLUTIONS AND RECOMMENDATIONS

The variety of QSPR/QSAR models developed with the CORAL software indicates that this approach can be an efficient tool of the QSPR/QSAR analyses. However, it is to be noted, there are problems related to utilization of the CORAL software.

Problem 1: The involvement of combinations of two and three SMILES atoms in building up a model can lead to appearance of SMILES attributes (Table 1) with unclear physical meaning, e.g.:

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

\...1.....

=...2.....

=...1...(...

[.../.....

[...=.....

Problem 2: The approach is able to detect three-atom fragments as basis for the mechanistic interpretation. In reality, an active fragment can involve larger number of atoms.

Problem 3: If all SMILES contain certain, an indeed active fragment, this fragment cannot be recognized by the CORAL software as an active one.

Problem 4: If an active fragment is rare, the fragment cannot be recognized by CORAL software as an active one.

However, a preliminary analysis and classification (filtration) of dataset can help to solve these problems, at least partially.

FUTURE RESEARCH DIRECTIONS

The recently published works have shown, that the CORAL software can be a tool of QSPR/QSAR analyses for other endpoints in addition to results listed in Table 3 (Singh and Gupta, 2014; Achary, 2014, pp. 73-90; Worachartcheewan, Nantasenamat, Isarankura-Na-Ayudhya, &Prachayasittikul, 2014).

QSAR Prediction Reporting Format (QPRF) (version 1.1, May 2008) available on the Internet http://ihcp.jrc.ec.europa.eu/our_labs/predictive_toxicology/qsar_tools/qrf/QPRF_version_1%201_DEREK_SS.pdf is multi-component system. Most likely, the system can be useful and attractive for the case of standardized situations. However, in order to be widely applied as a research tool, it is preferable to use OECD principles together with the following probabilistic rules:

1. A QSPR/QSAR model should be represented for a group of splits of data into the "visible" training set (compounds involved in the modeling process) and the "invisible" test set (compounds which are not involved in the modeling process);
2. The splits should be random and different (they should be far from identity). The Monte Carlo technique in general, and the CORAL software in particular, can be basis for the QSAR Prediction Reporting Format, at least for the first steps of the QSPR/QSAR analyses of ecologic, pharmacologic and other systems of endpoints which are characterized by high complexity and wide prevalence.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

CONCLUSION

The CORAL software can be considered as an universal tool for the QSPR/QSAR analysis. As all software packages the CORAL has certain advantages and disadvantages. The software provides the following options:

1. QSPR/QSAR analysis for data represented in format described at http://www.insilico.eu/coral;
2. The analyses should be carried out using a number of various data splits;
3. The software can be involved in build up models for phenomena characterized by eclectic information. Thus, the most important advantages of the approach are:
   a. The possibility of analysis of a QSPR/QSAR as a random event (not unique action); and
   b. The possibility of the translation of eclectic information into endpoint prediction.

ACKNOWLEDGMENT

We thank the EC project NANOPUZZLES (Project Reference: 309837). D.L. and J.L. acknowledge support from the National Science Foundation (NSF/CREST HRD-0833178). E.A.C., D.E.B. and P.R.D. thank for the financial support provided by the National Research Council of Argentina (CONICET) PIP11220100100151 project and to Ministerio de Ciencia, Tecnología e Innovación "Productiva for the electronic library facilities."

REFERENCES

Achary, P. G. R. (2014). Simplified molecular input line entry system-based optimal descriptors: QSAR modelling for voltage-gated potassium channel subunit Kv7.2. *SAR and QSAR in Environmental Research*, *25*(1), 73–90. doi:10.1080/1062936X.2013.842930 PMID:24580100

Afantitis, A., Melagraki, G., Sarimveis, H., Koutentis, P. A., Markopoulos, J., & Igglessi-Markopoulou, O. (2006). Prediction of intrinsic viscosity in polymer–solvent combinations using a QSPR model. *Polymer*, *47*(9), 3240–3248. doi:10.1016/j.polymer.2006.02.060

Atanasova, M., Ilieva, S., & Galabov, B. (2007). QSAR analysis of 1,4-dihydro-4-oxo-1-(2-thiazolyl)-1,8-naphthyridines with anticancer activity. *European Journal of Medicinal Chemistry*, *42*(9), 1184–1192. doi:10.1016/j.ejmech.2007.01.029 PMID:17408810

Balaban, A. T. (1982). Highly discriminating distance-based topological index. *Chemical Physics Letters*, *89*(5), 399–404. doi:10.1016/0009-2614(82)80009-2

Bonchev, D. (1995). Topological order in molecules 1. Molecular branching revisited. *Journal of Molecular Structure THEOCHEM*, *336*(2-3), 137–156. doi:10.1016/0166-1280(94)04081-3

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Bonchev, D., & Trinajsti, N. (1977). Information theory, distance matrix, and molecular branching. *The Journal of Chemical Physics*, *67*(10), 4517–4533. doi:10.1063/1.434593

Castro, E. A., Tueros, M., & Toropov, A. A. (2000). Maximum topological distances based indices as molecular descriptors for QSPR 2 — Application to aromatic hydrocarbons. *Computers & Chemistry*, *24*(5), 571–576. doi:10.1016/S0097-8485(99)00095-9 PMID:10890366

Coi, A., Massarelli, I., Murgia, L., Saraceno, M., Calderone, V., & Bianuccia, A. M. (2006). Prediction of hERG potassium channel affinity by the CODESSA approach. *Bioorganic & Medicinal Chemistry*, *14*(9), 3153–3159. doi:10.1016/j.bmc.2005.12.030 PMID:16426850

Cox, D. S., Gao, H., Raje, S., Scott, K. R., & Eddington, N. D. (2001a). Enhancing the permeation of marker compounds and enaminone anticonvulsants across Caco-2 monolayers by modulating tight junctions using zonula occludens toxin. *European Journal of Pharmaceutics and Biopharmaceutics*, *52*(2), 145–150. doi:10.1016/S0939-6411(01)00172-2 PMID:11522479

Cox, D. S., Scott, K. R., Gao, H., Raje, S., & Eddington, N. D. (2001b). Influence of multidrug resistance (MDR) proteins at the blood-brain barrier on the transport and brain distribution of enaminone anticonvulsants. *Journal of Pharmaceutical Sciences*, *90*(10), 1540–1552. doi:10.1002/jps.1104 PMID:11745712

Cronin, M. T. D., Aptula, A. O., Judith, C., Duffy, J. C., Netzeva, T. I., & Rowe, P. H. et al. (2002). Comparative assessment of methods to develop QSARs for the prediction of the toxicity of phenols to Tetrahymena pyriformis. *Chemosphere*, *49*(10), 1201–1221. doi:10.1016/S0045-6535(02)00508-8 PMID:12489717

da Silva Junkes, B., Arruda, A. C. S., Yunes, R. A., Porto, L. C., & Heinzen, V. E. F. (2005). Semi-empirical topological index: A tool for QSPR/QSAR studies. *Journal of Molecular Modeling*, *11*(2), 128–134. doi:10.1007/s00894-004-0231-8 PMID:15723207

de Melo, E. B. (2012). A new quantitative structure–property relationship model to predict bioconcentration factors of polychlorinated biphenyls (PCBs) in fishes using E-state index and topological descriptors. *Ecotoxicology and Environmental Safety*, *75*, 213–222. doi:10.1016/j.ecoenv.2011.08.026 PMID:21959189

Diudea, M. V., & Gutman, I. (1998). Wiener-type topological indices. *Croatica Chemica Acta*, *71*(1), 21–51.

Duchowicz, P. R., Castro, E. A., Fernandez, F. M., & Gonzalez, M. P. (2005). A new search algorithm for QSPR/QSAR theories: Normal boiling points of some organic molecules. *Chemical Physics Letters*, *412*(4-6), 376–380. doi:10.1016/j.cplett.2005.07.016

Duchowicz, P. R., Comelli, N. C., Ortiz, E. V., & Castro, E. A. (2012). QSAR study for carcinogenicity in a large set of organic compounds. *Current Drug Safety*, *7*, 282–288. doi:10.2174/157488612804096623 PMID:23062240

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Durdagi, S., Mavromoustakos, T., & Papadopoulos, M. G. (2008). 3D QSAR CoMFA/CoMSIA, molecular docking and molecular dynamics studies of fullerene-based HIV-1 PR inhibitors. *Bioorganic & Medicinal Chemistry Letters*, *18*(23), 6283–6289. doi:10.1016/j.bmcl.2008.09.107 PMID:18951793

Edafiogho, I. O., Ananthalakshmi, K. V., & Kombian, S. B. (2006). Anticonvulsant evaluation and mechanism of action of benzylamino enaminones. *Bioorganic & Medicinal Chemistry*, *14*(15), 5266–5272. doi:10.1016/j.bmc.2006.03.049 PMID:16621576

Eddington, N. D., Cox, D. S., Khurana, M., Salama, N. N., Stables, J. P., & Harrison, S. J. et al. (2003). Synthesis and anticonvulsant activity of enaminones part 7. Synthesis and anticonvulsant evaluation of ethyl 4-[(substituted phenyl)amino]-6-methyl-2-oxocyclohex-3-ene-1-carboxylates and their corresponding 5-methylcyclohex-2-enone derivatives. *European Journal of Medicinal Chemistry*, *38*(1), 49–64. doi:10.1016/S0223-5234(02)00006-5 PMID:12593916

Eddington, N. D., Cox, D. S., Roberts, R. R., Butcher, R. J., Edafiogho, I. O., & Stables, J. P. et al. (2002). Synthesis and anticonvulsant activity of enaminones. 4. Investigations on isoxazole derivatives. *European Journal of Medicinal Chemistry*, *37*(8), 635–648. doi:10.1016/S0223-5234(02)01377-6 PMID:12161061

Fouchécourt, M.-O., Béliveau, M., & Krishnan, K. (2001). Quantitative structure-pharmacokinetic relationship modeling. *The Science of the Total Environment*, *274*(1-3), 125–135. doi:10.1016/S0048-9697(01)00743-4 PMID:11453289

García, J., Duchowicz, P. R., Rozas, M. F., Caram, J. A., Mirífico, M. V., Fernández, F. M., & Castro, E. A. (2011). A comparative QSAR on 1,2,5-thiadiazolidin-3-one 1,1-dioxide compounds as selective inhibitors of human serine proteinases. *Journal of Molecular Graphics & Modelling*, *31*, 10–19. doi:10.1016/j.jmgm.2011.07.007 PMID:21908217

Garro Martinez, J. C., Duchowicz, P. R., Estrada, M. R., Zamarbide, G. N., & Castro, E. A. (2011). QSAR study and molecular design of open-chain enaminones as anticonvulsant agents. *International Journal of Molecular Sciences*, *12*(12), 9354–9368. doi:10.3390/ijms12129354 PMID:22272137

Geçen, N., Sarıpınar, E., Yanmaz, E., & Şahin, K. (2012). Application of electron conformational-genetic algorithm approach to 1,4-dihydropyridines as calcium channel antagonists: Pharmacophore identification and bioactivity prediction. *Journal of Molecular Modeling*, *18*(1), 65–82. doi:10.1007/s00894-011-1024-5 PMID:21451963

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

González, M. P., Terán, C., Teijeira, M., & González-Moa, M. J. (2005). GETAWAY descriptors to predicting A2A adenosine receptors agonists. *European Journal of Medicinal Chemistry*, *40*(11), 1080–1086. doi:10.1016/j.ejmech.2005.04.014 PMID:16006015

Gutman, I. (1988). Topological analysis of Eigenvalues of the adjacency matrices in graph theory: A difficulty with the concept of internal connectivity. *Chemical Physics Letters*, *148*(1), 93–94. doi:10.1016/0009-2614(88)87267-1

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Gutman, I. (1997). Permanents of adjacency matrices and their dependence on molecular structure. *Polycyclic Aromatic Compounds*, *12*(4), 281–287. doi:10.1080/10406639808233845

Hammett, L. P. (1935). Some relations between reaction rates and equilibrium constants. *Chemical Reviews*, *17*(1), 125–136. doi:10.1021/cr60056a010

Helguera, A. M., Cordeiro, M. N. D. S., Pérez, M. A. C., Combes, R. D., & González, M. P. (2008). QSAR modeling of the rodent carcinogenicity of nitrocompounds. *Bioorganic & Medicinal Chemistry*, *16*(6), 3395–3407. doi:10.1016/j.bmc.2007.11.029 PMID:18295489

Hemmateenejad, B. (2004). Optimal QSAR analysis of the carcinogenic activity of drugs by correlation ranking and genetic algorithm-based PCR. *Journal of Chemometrics*, *18*(11), 475–485. doi:10.1002/cem.891

Huuskonen, J. (2000). Estimation of aqueous solubility for a diverse set of organic compounds based on molecular topology. *Journal of Chemical Information and Computer Sciences*, *40*(3), 773–777. doi:10.1021/ci9901338 PMID:10850781

Ibezim, E., Duchowicz, P. R., Ortiz, E. V., & Castro, E. A. (2012). QSAR on aryl-piperazine derivatives with activity on malaria. *Chemometrics and Intelligent Laboratory Systems*, *110*(1), 81–88. doi:10.1016/j.chemolab.2011.10.002

Ivanciuc, O., Ivanciuc, T., & Balaban, A. T. (1998). Quantitative structure-property relationship study of normal boiling points for halogen-/oxygen-/sulfur-containing organic compounds using the CODESSA program. *Tetrahedron*, *54*(31), 9129–9142. doi:10.1016/S0040-4020(98)00550-X

Kar, S., & Roy, K. (2010). QSAR modeling of toxicity of diverse organic chemicals to Daphnia magna using 2D and 3D descriptors. *Journal of Hazardous Materials*, *177*(1-3), 344–351. doi:10.1016/j.jhazmat.2009.12.038 PMID:20045248

Kar, S., & Roy, K. (2012). First report on development of quantitative interspecies structure–carcinogenicity relationship models and exploring discriminatory features for rodent carcinogenicity of diverse organic chemicals using OECD guidelines. *Chemosphere*, *87*(4), 339–355. doi:10.1016/j.chemosphere.2011.12.019 PMID:22225702

Kar, S. K., & Roy, K. (2011). Development and validation of a robust QSAR model for prediction of carcinogenicity of drugs. *Indian Journal of Biochemistry & Biophysics*, *48*, 111–122. PMID:21682143

Kruhlak, N. L., Contrera, J. F., Daniel Benz, R., & Matthews, E. J. (2007). Progress in QSAR toxicity screening of pharmaceutical impurities and other FDA regulated products. *Advanced Drug Delivery Reviews*, *59*(1), 43–55. doi:10.1016/j.addr.2006.10.008 PMID:17229485

Kušić, H., Rasulev, B., Leszczynska, D., Leszczynski, J., & Koprivanac, N. (2009). Prediction of rate constants for radical degradation of aromatic pollutants in water matrix: A QSAR study. *Chemosphere*, *75*(8), 1128–1134. doi:10.1016/j.chemosphere.2009.01.019 PMID:19201442

Leonard, J. T., & Roy, K. (2006). On selection of training and test sets for the development of predictive QSAR models. *QSAR & Combinatorial Science*, *25*(3), 235–251. doi:10.1002/qsar.200510161

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Martinez-Esparza, J., Oficialdegui, A. M., Perez-Silanes, S., Heras, B. B., Orus, L., & Palop, J. A. et al. (2001a). New 1-Aryl-3-(4-arylpiperazin-1-yl)propane derivatives, with dual action at 5-HT1A serotonin receptors and serotonin transporter, as a new class of antidepressants. *Journal of Medicinal Chemistry*, *44*(3), 418–428. doi:10.1021/jm001059j PMID:11462981

Martinez-Esparza, J., Perez-Silanes, S., Oficialdegui, A. M., Heras, B., Orus, L., & Villanueva, H. et al. (2001b). New 3-[4-(aryl)piperazin- 1-yl]-1-(benzo[b]thiophen-3-yl)propane derivatives with dual action at 5-HT1A serotonin receptors and serotonin transporter as a new class of antidepressants. *European Journal of Medicinal Chemistry*, *36*(1), 55–61. doi:10.1016/S0223-5234(00)01198-3 PMID:11231049

Melagraki, G., Afantitis, A., Sarimveis, H., Igglessi-Markopouloua, O., & Supuran, C. T. (2006). QSAR study on para-substituted aromatic sulfonamides as carbonic anhydrase II inhibitors using topological information indices. *Bioorganic & Medicinal Chemistry*, *14*(4), 1108–1114. doi:10.1016/j.bmc.2005.09.038 PMID:16213737

Molyneaux, C.-A., Krugliak, M., Ginsburg, H., & Chibale, K. (2005). Arylpiperazines displaying preferential potency against chloroquine-resistant strains of the malaria parasite Plasmodium falciparum. *Biochemical Pharmacology*, *71*(1-2), 61–68. doi:10.1016/j.bcp.2005.10.023 PMID:16298348

Mullen, L. M. A., Duchowicz, P. R., & Castro, E. A. (2011). QSAR treatment on a new class of triphenylmethyl-containing compounds as potent anticancer agents. *Chemometrics and Intelligent Laboratory Systems*, *107*(2), 269–275. doi:10.1016/j.chemolab.2011.04.011

Nesmerak, K., Nemec, I., Sticha, M., Waisser, K., & Palat, K. (2005). Quantitative structure–property relationships of new benzoxazines and their electrooxidation as a model of metabolic degradation. *Electrochimica Acta*, *50*(6), 1431–1437. doi:10.1016/j.electacta.2004.08.031

Nesmerak, K., Toropov, A. A., & Toropova, A. (2014). SMILES-based quantitative structure–retention relationships for RP-HPLC of 1-phenyl-5-benzylsulfanyltetrazoles. *Structural Chemistry*, *25*(1), 311–317. doi:10.1007/s11224-013-0293-2

Nesmerak, K., Toropov, A. A., Toropova, A. A., Kohoutova, P., & Waisser, K. (2013). SMILES-based quantitative structureeproperty relationships for half-wave potential of N-benzylsalicylthioamides. *European Journal of Medicinal Chemistry*, *67*, 111–114. doi:10.1016/j.ejmech.2013.05.031 PMID:23850571

Nikolic, S., Kovacevic, G., Milicevic, A., & Trinajstic, N. (2003). The Zagreb indices 30 years after. *Croatica Chemica Acta*, *76*(2), 113–124.

Oberg, T. (2005). A QSAR for the hydroxylradicalreaction rate constant: Validation, domain of application, and prediction. *Atmospheric Environment*, *39*(12), 2189–2200. doi:10.1016/j.atmosenv.2005.01.007

OECD. (2007). *Quantitative structure-activity relationships project [(Q)SARs]*. Available online: http://search.oecd.org/officialdocuments/displaydocumentpdf/?doclanguage=en&cote=env/jm/mono(2007)2

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Ojha, P. K., & Roy, K. (2011). Comparative QSARs for antimalarial endochins: Importance of descriptor-thinning and noise reduction prior to feature selection. *Chemometrics and Intelligent Laboratory Systems*, *109*(2), 146–161. doi:10.1016/j.chemolab.2011.08.007

Palchaudhuri, R., & Hergenrother, P. J. (2008). Triphenylmethylamides (TPMAs): Structure-activity relationship of compounds that induce apoptosis in melanoma cells. *Bioorganic & Medicinal Chemistry Letters*, *18*(22), 5888–5891. doi:10.1016/j.bmcl.2008.07.128 PMID:18710803

Palchaudhuri, R., Nesterenko, V., & Hergenrother, P. J. (2008). The complex role of the triphenylmethyl motif in anticancer compounds. *Journal of the American Chemical Society*, *130*(31), 10274–10281. doi:10.1021/ja8020999 PMID:18611022

Pan, Y., Jiang, J., & Wang, Z. (2007). Quantitative structure–property relationship studies for predicting flash points of alkanes using group bond contribution method with back-propagation neural network. *Journal of Hazardous Materials*, *147*(1-2), 424–430. doi:10.1016/j.jhazmat.2007.01.025 PMID:17292543

Papa, E., van derWal, L., Arnot, J. A., & Gramatica, P. (2014). Metabolic biotransformation half-lives in fish: QSAR modeling and consensus analysis. *The Science of the Total Environment*, *470–471*, 1040–1046. doi:10.1016/j.scitotenv.2013.10.068 PMID:24239825

Porto, L. C., Souza, E. S., Junkes, B. S., Yunes, R. A., & Heinzen, V. E. F. (2008). Semi-empirical topological index: Development of QSPR/QSRR and optimization for alkylbenzenes. *Talanta*, *76*(2), 407–412. doi:10.1016/j.talanta.2008.03.023 PMID:18585298

Putz, M. V., Ionascu, C., Putz, A.-M., & Ostafe, V. (2011). Alert-QSAR. Implications for electrophilic theory of chemical carcinogenesis. *International Journal of Molecular Sciences*, *12*(12), 5098–5134. doi:10.3390/ijms12085098 PMID:21954348

Randić, M. (1975). Characterization of molecular branching. *Journal of the American Chemical Society*, *69*(23), 6609–6615. doi:10.1021/ja00856a001

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

Randic, M., & Basak, S. C. (1999). Optimal molecular descriptors based on weighted path numbers. *Journal of Chemical Information and Computer Sciences*, *39*(2), 261–266. doi:10.1021/ci9800763

Randic, M., & Basak, S. C. (2001). New descriptor for structure-property and structure-activity correlations. *Journal of Chemical Information and Computer Sciences*, *41*(3), 650–656. doi:10.1021/ci000116e PMID:11410042

Randic, M., DeAlba, L. M., & Harris, F. E. (1998). Graphs with the same detour matrix. *Croatica Chemica Acta*, *71*(1), 53–68.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Randic, M., & Pompe, M. (2001). The variable connectivity index 1xf versus the traditional molecular descriptors: A comparative study of 1xf against descriptors of CODESSA. *Journal of Chemical Information and Computer Sciences*, *41*(3), 631–638. doi:10.1021/ci000119r PMID:11410040

Ren, B. (2003). Atom-type-based AI topological descriptors for quantitative structure–retention index correlations of aldehydes and ketones. *Chemometrics and Intelligent Laboratory Systems*, *66*(1), 29–39. doi:10.1016/S0169-7439(03)00004-2

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

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

Singh, K. P., & Gupta, S. (2014). Nano-QSAR modeling for predicting biological activity of diverse nanomaterials. *Royal Society of Chemistry Advances*, *4*(26), 13215–13230.

Talevi, A., Bellera, C. L., Ianni, M. D., Duchowicz, P. R., Bruno-Blanch, L. E., & Castro, E. A. (2012). An integrated drug development approach. Applying topological descriptors. *Current Computer-aided Drug Design*, *8*, 172–181. doi:10.2174/157340912801619076 PMID:22734704

Taskinen, J., & Yliruusi, J. (2003). Prediction of physicochemical properties based on neural network modelling. *Advanced Drug Delivery Reviews*, *55*(9), 1163–1183. doi:10.1016/S0169-409X(03)00117-0 PMID:12954197

Tetko, I. V., Tanchuk, V. Yu., Kasheva, T. N., & Villa, A. E. P. (2001). Estimation of aqueous solubility of chemical compounds using E-state indices. *Journal of Chemical Information and Computer Sciences*, *41*(6), 1488–1493. doi:10.1021/ci000392t PMID:11749573

Toropov, A. A., Nesmerak, K., Raska, I. Jr, Waisser, K., & Palat, K. (2006). QSPR modeling of the half-wave potentials of benzoxazines by optimal descriptors calculated with the SMILES. *Computational Biology and Chemistry*, *30*(6), 434–437. doi:10.1016/j.compbiolchem.2006.09.003 PMID:17092778

Toropov, A. A., Rasulev, B. F., & Leszczynski, J. (2008). QSAR modeling of acute toxicity by balance of correlations. *Bioorganic & Medicinal Chemistry*, *16*(11), 5999–6008. doi:10.1016/j.bmc.2008.04.055 PMID:18482841

Toropov, A. A., & Toropova, A. P. (2001). Prediction of heteroaromatic amine mutagenicity by means of correlation weighting of atomic orbital graphs of local invariants. *Journal of Molecular Structure THEOCHEM*, *538*(1-3), 287–293. doi:10.1016/S0166-1280(00)00713-2

Toropov, A. A., & Toropova, A. P. (2002). QSAR modeling of toxicity on optimization of correlation weights of Morgan extended connectivity. *Journal of Molecular Structure THEOCHEM*, *578*(1-3), 129–134. doi:10.1016/S0166-1280(01)00695-9

Toropov, A. A., Toropova, A. P., & Benfenati, E. (2010b). SMILES-based optimal descriptors: QSAR modeling of carcinogenicity by balance of correlations with ideal slopes. *European Journal of Medicinal Chemistry*, *45*(9), 3581–3587. doi:10.1016/j.ejmech.2010.05.002 PMID:20570021

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Toropov, A. A., Toropova, A. P., Benfenati, E., Gini, G., & Fanelli, R. (2013b). The definition of the molecular structure for potential anti-malaria agents by the Monte Carlo method. *Structural Chemistry*, *24*(4), 1369–1381. doi:10.1007/s11224-012-0180-2

Toropov, A. A., Toropova, A. P., Benfenati, E., Gini, G., Leszczynska, D., & Leszczynski, J. (2010c). QSAR analysis of 1,4-dihydro-4-oxo-1-(2-thiazolyl)-1,8-naphthyridines exhibiting anticancer activity by optimal SMILES-based descriptors. *Journal of Mathematical Chemistry*, *47*(2), 647–666. doi:10.1007/s10910-009-9593-6

Toropov, A. A., Toropova, A. P., Benfenati, E., Gini, G., Leszczynska, D., & Leszczynski, J. (2013a). CORAL: QSPR model of water solubility based on local and global SMILES attributes. *Chemosphere*, *90*(2), 877–880. doi:10.1016/j.chemosphere.2012.07.035 PMID:22921649

Toropov, A. A., Toropova, A. P., Benfenati, E., Leszczynska, D., & Leszczynski, J. (2010d). SMILES-based optimal descriptors: QSAR analysis of fullerene-based HIV-1 PR inhibitors by means of balance of correlations. *Journal of Computational Chemistry*, *31*, 381–392. PMID:19479738

Toropov, A. A., Toropova, A. P., Raska, I. Jr, & Benfenati, E. (2010a). QSPR modeling of octanol/water partition coefficient of antineoplastic agents by balance of correlation. *European Journal of Medicinal Chemistry*, *45*(4), 1639–1647. doi:10.1016/j.ejmech.2009.12.043 PMID:20106561

Toropov, A. A., Toropova, A. P., Raska, I. Jr, Benfenati, E., & Gini, G. (2012b). QSAR modeling of endpoints for peptides which is based on representation of the molecular structure by a sequence of amino acids. *Structural Chemistry*, *23*(6), 1891–1904. doi:10.1007/s11224-012-9995-0

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

Toropov, A. A., Toropova, A. P., Rasulev, B. F., Benfenati, E., Gini, G., Leszczynska, D., & Leszczynski, J. (2012a). Coral: QSPR modeling of rate constants of reactions between organic aromatic pollutants and hydroxyl radical. *Journal of Computational Chemistry*, *33*(23), 1902–1906. doi:10.1002/jcc.23022 PMID:22641453

Toropova, A. P., & Toropov, A. A. (2013). Optimal descriptor as a translator of eclectic information into the prediction of membrane damage by means of various TiO_2 nanoparticles. *Chemosphere*, *93*(10), 2650–2655. doi:10.1016/j.chemosphere.2013.09.089 PMID:24161577

Toropova, A. P., & Toropov, A. A. (2014). CORAL software: Prediction of carcinogenicity of drugs by means of the Monte Carlo method. *European Journal of Pharmaceutical Sciences*, *52*, 21–25. doi:10.1016/j.ejps.2013.10.005 PMID:24514451

Toropova, A. P., Toropov, A. A., Benfenati, E., Gini, G., Leszczynska, D., & Leszczynski, J. (2011a). CORAL: Quantitative structure-activity relationship models for estimating toxicity of organic compounds in rats. *Journal of Computational Chemistry*, *32*(12), 2727–2733. doi:10.1002/jcc.21848 PMID:21656789

Toropova, A. P., Toropov, A. A., Benfenati, E., Gini, G., Leszczynska, D., & Leszczynski, J. (2011b). CORAL: QSPR models for solubility of [C60] and [C70] fullerene derivatives. *Molecular Diversity*, *15*(1), 249–256. doi:10.1007/s11030-010-9245-6 PMID:20349134

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Toropova, A. P., Toropov, A. A., Benfenati, E., Gini, G., Leszczynska, D., & Leszczynski, J. (2012b). CORAL: Models of toxicity of binary mixtures. *Chemometrics and Intelligent Laboratory Systems*, *119*, 39–43. doi:10.1016/j.chemolab.2012.10.001

Toropova, A. P., Toropov, A. A., Benfenati, E., Leszczynska, D., & Leszczynski, J. (2010). QSAR modeling of measured binding affinity for fullerene-based HIV-1 PR inhibitors by CORAL. *Journal of Mathematical Chemistry*, *48*(4), 959–987. doi:10.1007/s10910-010-9719-x

Toropova, A. P., Toropov, A. A., Martyanov, S. E., Benfenati, E., Gini, G., Leszczynska, D., & Leszczynski, J. (2012a). CORAL: QSAR modeling of toxicity of organic chemicals towards *Daphnia magna*. *Chemometrics and Intelligent Laboratory Systems*, *110*(1), 177–181. doi:10.1016/j.chemolab.2011.10.005

Troshin, P. A., Hoppe, H., Renz, J., Egginger, M., Mayorova, J. Yu., & Goryachev, A. E. et al. (2009). Material solubility-photovoltaic performance relationship in the design of novel fullerene derivatives for bulk heterojunction solar cells. *Advanced Functional Materials*, *19*(5), 779–788. doi:10.1002/adfm.200801189

Turabekova, M. A., Rasulev, B. F., Dzhakhangirov, F. N., Toropov, A. A., Leszczynska, D., & Leszczynski, J. (2014). *Aconitum* and *Delphinium* diterpenoid alkaloids of local anesthetic activity. Comparative QSAR analysis based on GA-MLRA/PLS and optimal descriptors approach. *Journal of Environmental Science and Health*, *32*(Part C), 213–238. doi:10.1080/10590501.2014.938886 PMID:25226219

Vamecq, J., Lambert, D., Poupaert, J. H., Masereel, B., & Stables, J. P. (1998). Anticonvulsant activity and interactions with neuronal voltage-dependent sodium channel of analogues of ameltolide. *Journal of Medicinal Chemistry*, *41*(18), 3307–3313. doi:10.1021/jm9608772 PMID:9719582

Veselinovic, A. M., Milosavljevic, J. B., Toropov, A., & Nikolic, G. M. (2013a). SMILES-based QSAR model for arylpiperazines as high-affinity 5-HT1A receptor ligands using CORAL. *European Journal of Pharmaceutical Sciences*, *48*(3), 532–541. doi:10.1016/j.ejps.2012.12.021 PMID:23287365

Veselinovic, A. M., Milosavljevic, J. B., Toropov, A., & Nikolic, G. M. (2013b). SMILES-Based QSAR Models for the Calcium Channel-Antagonistic Effect of 1,4-Dihydropyridines. *Archiv der Pharmazie Chemistry in Life Sciences*, *346*(2), 134–139. doi:10.1002/ardp.201200373 PMID:23280520

Vracko, M. (1997). A study of structure-carcinogenic potency relationship with artificial neural networks. The using of descriptors related to geometrical and electronic structures. *Journal of Chemical Information and Computer Sciences*, *37*(6), 1037–1043. doi:10.1021/ci970231y

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. (1990) SMILES. 3. Depict. Graphical depiction of chemical structures. *Journal of Chemical Information and Computer Sciences, 30*(3), 237-243.

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

Wiener, H. (1948). Vapor pressure-temperature relationships among the branched paraffin hydrocarbons. *The Journal of Physical and Colloid Chemistry*, *52*(2), 425–430. doi:10.1021/j150458a014 PMID:18906414

Wishart, D. S., Knox, C., Guo, A. C., Cheng, D., Shrivastava, S., & Tzur, D. et al. (2008). DrugBank: A knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Research*, *36*(Database), 901–906. doi:10.1093/nar/gkm958 PMID:18048412

Worachartcheewan, A., Nantasenamat, C., Isarankura-Na-Ayudhya, C., & Prachayasittikul, V. (2014). QSAR study of H1N1 neuraminidase inhibitors from influenza a virus. *Letters in Drug Design & Discovery*, *11*(4), 420–427. doi:10.2174/15701808113106660085

Yan, A., & Gasteiger, J. (2003). Prediction of aqueous solubility of organic compounds based on a 3D structure representation. *Journal of Chemical Information and Computer Sciences*, *43*(2), 429–434. doi:10.1021/ci025590u PMID:12653505

Zhang, L., Zhou, P., Yang, F., & Wang, Z. (2007). Computer-based QSARs for predicting mixture toxicity of benzene and its derivatives. *Chemosphere*, *67*(2), 396–401. doi:10.1016/j.chemosphere.2006.09.018 PMID:17184822

KEY TERMS AND DEFINITIONS

Correlation Weight: A coefficient which should be added to optimal descriptor if a given SMILES attribute or graph invariant takes place in the given SMILES or in the given molecular graph.

Global SMILES Attribute: A characteristic of molecules in whole, e.g. the presence of nitrogen together with oxygen, presence of double bonds and cycles, the presence of nitrogen and sulphur and absence of oxygen, and others (Table 1).

Graph of Atomic Orbitals (GAO): Molecular graph where vertexes are the representation of atomic orbitals (e.g. 1s2, 2p3, 3d7, etc.).

Hydrogen-Filled Graph (HFG): Molecular graph where presence of hydrogen atoms is taken into account.

Hydrogen-Suppressed Graph (HSG): Molecular graph where presence of hydrogen atoms is ignored.

Local SMILES Attribute: Fragment of SMILES line which contains one ('C', 'N', '=', etc.) or two symbols ('Cl', 'Br', '@@', etc.) which cannot be examined separately.

Monte Carlo Optimization: The calculation of the numerical data on the correlation weights which give maximal value of the correlationn coefficient between an endpoint and the optimal descriptor.

Optimal Descriptor: A descriptor which is calculated for a substances with simplified molecular input-line entry system (SMILES) or with molecular graph using correlation weights for SMILES attributes or for graph invariant (vertex degree, valence shells, etc.).

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Compilation of 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

Aboye, T. L., Sobhia, M. E., & Bharatam, P. V. (2004). 3D-QSAR studies of pyruvate dehydrogenase kinase inhibitors based on a divide and conquer strategy. *Bioorganic & Medicinal Chemistry*, *12*(10), 2709–2715. doi:10.1016/j.bmc.2004.03.019 PMID:15110852

Abraham, M. H. (1994). New solute descriptors for linear free energy relationships and quantitative structure-activity relationships. In Quantitative treatments of solute/solvent interactions. Elsevier.

Abraham, M. H., Chadha, H. S., Dixon, J. P., Rafols, C., & Treiner, C. (1995). Hydrogen bonding. Part 40. Factors that influence the distribution of solutes between water and sodium dodecylsulfate micelles. *Journal of the Chemical Society, Perkin Transactions*, *2*(5), 887–894. doi:10.1039/p29950000887

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

Accelrys Inc. (2011). Discovery studio 3.0. San Diego: Author.

Achary, P. G. R. (2014). Simplified molecular input line entry system-based optimal descriptors: QSAR modelling for voltage-gated potassium channel subunit Kv7.2. *SAR and QSAR in Environmental Research*, *25*(1), 73–90. doi:10.1080/1062936X.2013.842930 PMID:24580100

Adhikari, N., Halder, A. K., Mondal, C., & Jha, T. (2013). Exploring structural requirements of aurone derivatives as antimalarials by validated DFT-based QSAR, HQSAR, and COMFA–COMSIA approach. *Medicinal Chemistry Research*, *22*(12), 6029–6045. doi:10.1007/s00044-013-0590-8

Adhikari, N., Halder, A. K., Mondal, C., & Jha, T. (2014). Structural findings of quinolone carboxylic acids in cytotoxic, antiviral, and anti-HIV-1 integrase activity through validated comparative molecular modeling studies. *Medicinal Chemistry Research*, *23*(6), 3096–3127. doi:10.1007/s00044-013-0897-5

Adhikari, N., Maiti, M. K., & Jha, T. (2010). Predictive comparative QSAR modelling of (phenylpiperazinylalkyl) 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

Afantitis, A., Melagraki, G., Sarimveis, H., Koutentis, P. A., Markopoulos, J., & Igglessi-Markopoulou, O. (2006). Prediction of intrinsic viscosity in polymer–solvent combinations using a QSPR model. *Polymer*, *47*(9), 3240–3248. doi:10.1016/j.polymer.2006.02.060

Agrafiotis, D. K. (2000). Multiobjective optimization of combinatorial libraries. *Molecular Diversity*, *5*(4), 209–230. doi:10.1023/A:1021320124615 PMID:12549673

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

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

AJCC Cancer Staging Manual. (2002). New York, NY: Springer Verlag.

Akama, H., Kanemaki, M., Yoshimura, M., Tsukihara, T., Kashiwagi, T., & Yoneyama, H. et al. (2004). Crystal structure of the drug discharge outer membrane protein, OprM, of *Pseudomonas aeruginosa* dual modes of membrane anchoring and occluded cavity end. *The Journal of Biological Chemistry*, *279*(51), 52816–52819. doi:10.1074/jbc.C400445200 PMID:15507433

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

Albert, A., Rubbo, S. D., & Goldacre, R. (1941). Correlation of basicity and antiseptic action in an acridine series. *Nature*, *147*(3724), 332–333. doi:10.1038/147332a0

Alekshun, M., & Nelson, M. (2004). *WO patent no. 2004062674*.

Alivisatos, P. (2004). The use of nanocrystals in biological detection. *Nature Biotechnology*, *22*(1), 47–52. doi:10.1038/nbt927 PMID:14704706

Allen, C., Maysinger, D., & Eisenberg, A. (1999). Nano-engineering block copolymer aggregates for drug delivery. *Colloids and Surfaces. B, Biointerfaces*, *16*(1-4), 3–27. doi:10.1016/S0927-7765(99)00058-2

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

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

Alter, M. J. (1993). The detection, transmission, and outcome of hepatitis C virus infection. *Infectious Agents and Disease*, *2*(3), 155–166. PMID:8173786

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

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

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

Ambrus, J. I., Kelso, M. J., Bremner, J. B., Ball, A. R., Casadei, G., & Lewis, K. (2008). Structure–activity relationships of 2-aryl-1*H*-indole inhibitors of the NorA efflux pump in *Staphylococcus aureus*. *Bioorganic & Medicinal Chemistry Letters*, *18*(15), 4294–4297. doi:10.1016/j.bmcl.2008.06.093 PMID:18632270

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

American Cancer Society Cancer Facts & Figures 2013. (2013). Atlanta, GA: American Cancer Society.

Amin, E. A., & Welsh, W. J. (2006). Highly predictive CoMFA and CoMSIA models for two series of stromelysin-1 (MMP-3) inhibitors elucidate S1‘ and S1−S2‘ binding modes. *Journal of Chemical Information and Modeling*, *46*(4), 1775–1783. doi:10.1021/ci060089d PMID:16859309

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

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

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

Antunes, J. E., Freitas, M. P., da Cunha, E. F. F., Ramalho, T. C., & Rittner, R. (2008). In silico prediction of novel phosphodiesterase type-5 inhibitors derived from Sildenafil, Vardenafil and Tadalafil. *Bioorganic & Medicinal Chemistry*, *16*(16), 7599–7606. doi:10.1016/j.bmc.2008.07.022 PMID:18656371

Aparna, V., Dineshkumar, K., Mohanalakshmi, N., Velmurugan, D., & Hopper, W. (2014). Identification of natural compound inhibitors for multidrug efflux pumps of *Escherichia coli* and *Pseudomonas aeruginosa* using *In Silico* high-throughput virtual screening and *In Vitro* validation. *PLoS ONE*, *9*(7), e101840. doi:10.1371/journal.pone.0101840 PMID:25025665

Aparna, V., Mohanalakshmi, N., Dineshkumar, K., & Hopper, W. (2014). Identification of inhibitors for RND efflux pump of *Pseudomonas aeruginosa* using structure-based pharmacophore modeling approach. *International Journal of Pharmacy & Pharmaceutical Sciences*, *6*(1).

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

Arning, J., Stolte, S., Böschen, A., Stock, F., Pitner, W.-R., & Welz-Biermann, U. et al. (2008). Qualitative and quantitative structure activity relationships for the inhibitory effects of cationic head groups, functionalised side chains and anions of ionic liquids on acetylcholinesterase. *Green Chemistry*, *10*(1), 47–58. doi:10.1039/b712109a

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

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

Aruoja, V., Dubourguier, H. C., Kasemets, K., & Kahru, A. (2009). Toxicity of nanoparticles of CuO, ZnO and TiO_2 to microalgae Pseudokirchneriella subcapitata. *The Science of the Total Environment*, *407*(4), 1461–1468. doi:10.1016/j.scitotenv.2008.10.053 PMID:19038417

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

Atal, C., Zutshi, U., & Rao, P. (1981). Scientific evidence on the role of Ayurvedic herbals on bioavailability of drugs. *Journal of Ethnopharmacology*, *4*(2), 229–232. doi:10.1016/0378-8741(81)90037-4 PMID:7311598

Atanasova, M., Ilieva, S., & Galabov, B. (2007). QSAR analysis of 1,4-dihydro-4-oxo-1-(2-thiazolyl)-1,8-naphthyridines with anticancer activity. *European Journal of Medicinal Chemistry*, *42*(9), 1184–1192. doi:10.1016/j.ejmech.2007.01.029 PMID:17408810

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

Avery, M. A., Alvim-Gaston, M., Rodrigues, C. R., Barreiro, E. J., Cohen, F. E., Sabnis, Y. A., & Woolfrey, J. R. (2002). Structure−activity relationships of the antimalarial agent artemisinin. 6. The development of predictive in vitro potency models using CoMFA and HQSAR methodologies. *Journal of Medicinal Chemistry*, *45*(2), 292–303. doi:10.1021/jm0100234 PMID:11784134

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

Ayers, P. W., & Levy, M. (2000). Perspective on "Density functional approach to the frontier-electron theory of chemical reactivity". *Theoretical Chemistry Accounts*, *103*(3-4), 353–360. doi:10.1007/s002149900093

Ayers, P. W., & Parr, R. G. (2000). Variational principles for describing chemical reactions: The Fukui function and chemical hardness revisited. *Journal of the American Chemical Society*, *122*(9), 2010–2018. doi:10.1021/ja9924039

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

Bachas, S., Eginton, C., Gunio, D., & Wade, H. (2011). Structural contributions to multidrug recognition in the multidrug resistance (MDR) gene regulator, BmrR. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(27), 11046–11051. doi:10.1073/pnas.1104850108 PMID:21690368

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

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. (2010). The use of QSAR and computational methods in drug design. In *Recent advances in QSAR studies* (pp. 261–282). Springer. doi:10.1007/978-1-4020-9783-6_9

Bakry, R., Vallant, R. M., Najam-ul-Haq, M., Rainer, M., Szabo, Z., Huck, C. W., & Bonn, G. K. (2007). Medicinal applications of fullerenes. *International Journal of Nanomedicine*, *2*(4), 639–649. PMID:18203430

Balaban, A. T. (1982). Highly discriminating distance based topological index. *Chemical Physics Letters*, *89*(5), 399–404. doi:10.1016/0009-2614(82)80009-2

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Balaban, A. T. (1997). From chemical topology to 3D geometry. *Journal of Chemical Information and Computer Sciences*, *37*(4), 645–650. doi:10.1021/ci960168x

Balaban, A. T. (1998). Topological and stereochemical molecular descriptors for databases useful in QSAR, similarity/dissimilarity and drug design. *SAR and QSAR in Environmental Research*, *8*(1-2), 1–21. doi:10.1080/10629369808033259

Ball, A. R., Casadei, G., Samosorn, S., Bremner, J. B., Ausubel, F. M., Moy, T. I., & Lewis, K. (2006). Conjugating berberine to a multidrug resistance pump inhibitor creates an effective antimicrobial. *ACS Chemical Biology*, *1*(9), 594–600. doi:10.1021/cb600238x PMID:17168555

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

Bao, Y., Huang, Q., Li, Y., Li, N., He, T., & Feng, C. (2012). Prediction of nitrobenzene toxicity to the algae (Scenedesmus obliguus) by quantitative structure–toxicity relationship (QSTR) models with quantum chemical descriptors. *Environmental Toxicology and Pharmacology*, *33*(1), 39–45. doi:10.1016/j.etap.2011.09.003 PMID:22088683

Bar-Ilan, O., Albrecht, R. M., Fako, V. E., & Furgeson, D. Y. (2009). Toxicity assessments of multisized gold and silver nanoparticles in zebrafish embryos. *Small*, *5*(16), 1897–1910. doi:10.1002/smll.200801716 PMID:19437466

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

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

Barnard, E. A. (1974). The peripheral nervous system. New York: Plenum Press.

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

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.

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

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

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 2/14/2023 7:16 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

Batterman, S., Chernyak, S., Gwynn, E., Cantonwine, D., Jia, C., Begnoche, L., & Hickey, J. P. (2007). Trends of brominated diphenyl ethers in fresh and archived Great Lakes fish (1979-2005). *Chemosphere*, *69*(3), 444–457. doi:10.1016/j.chemosphere.2007.04.066 PMID:17561231

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

Baun, A., Hartmann, N. B., Grieger, K., & Kusk, K. O. (2008). Ecotoxicity of engineered nanoparticles to aquatic invertebrates: A brief review and recommendations for future toxicity testing. *Ecotoxicology (London, England)*, *17*(5), 387–395. doi:10.1007/s10646-008-0208-y PMID:18425578

Bavro, V. N., Pietras, Z., Furnham, N., Pérez-Cano, L., Fernández-Recio, J., & Pei, X. Y. et al. (2008). Assembly and channel opening in a bacterial drug efflux machine. *Molecular Cell*, *30*(1), 114–121. doi:10.1016/j.molcel.2008.02.015 PMID:18406332

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

Bay, D. C., Rommens, K. L., & Turner, R. J. (2008). Small multidrug resistance proteins: A multidrug transporter family that continues to grow. *Biomembranes*, *1778*(9), 1814–1838. doi:10.1016/j.bbamem.2007.08.015

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

Bearden, A. P., & Schultz, T. W. (1998). Comparison of *Tetrahymena* and *Pimephales* toxicity based on mechanism of action. *SAR and QSAR in Environmental Research*, *9*(3-4), 127–153. doi:10.1080/10629369808039153 PMID:9933957

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

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

Bell, P. H., & Roblin, R. O. Jr. (1942). 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

Belsare, D. P., Pal, S. C., Kazi, A. A., Kankate, R. S., & Vanjari, S. S. (2010). Evaluation of antioxidant activity of chalcones and flavonoids. *International Journal of ChemTech Research*, *2*(2), 1080–1089.

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Bender, A., & Glen, R. C. (2004). Molecular similarity: A key technique in molecular informatics. *Journal of Organic and Biomolecular Chemistry*, *2*(22), 3204–3218. doi:10.1039/b409813g PMID:15534697

Benigni, R., & Giuliani, A. (2003). Putting the predictive toxicology challenge into perspective: Reflections on the results. *Bioinformatics (Oxford, England)*, *19*(10), 1194–1200. doi:10.1093/bioinformatics/btg099 PMID:12835261

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

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

Bergstrom, C. A. S., Wassvik, C. M., Norinder, U., Luthman, K., & Artursson, P. (2004). Global and local computational models for aqueous solubility prediction of drug-like molecules. *Journal of Chemical Information and Computer Sciences*, *44*(4), 1477–1488. doi:10.1021/ci049909h PMID:15272856

Berkowitz, M., & Parr, R. G. (1988). Molecular hardness and softness, local hardness and softness, hardness and softness kernels, and relations among these quantities. *The Journal of Chemical Physics*, *88*(4), 2554–2557. doi:10.1063/1.454034

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

Beyer, A., & Matthies, M. (2002). *Environmental long-range transport and persistence of substances. ELPOS 1.0.1. institute for environmental systems research*. Germany: University of Osnabruck.

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.

Bhardwaj, R. K., Glaeser, H., Becquemont, L., Klotz, U., Gupta, S. K., & Fromm, M. F. (2002). Piperine, a major constituent of black pepper, inhibits human P-glycoprotein and CYP3A4. *The Journal of Pharmacology and Experimental Therapeutics*, *302*(2), 645–650. doi:10.1124/jpet.102.034728 PMID:12130727

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

Bhongade, B. A., & Gadad, A. K. (2004). 3D-QSAR CoMFA/CoMSIA studies on urokinase plasminogen activator (uPA) inhibitors: A strategic design in novel anticancer agents. *Bioorganic & Medicinal Chemistry*, *12*(10), 2797–2805. doi:10.1016/j.bmc.2004.02.019 PMID:15110861

Bianco, A., Kostarelos, K., Partidos, C. D., & Prato, M. (2005). Biomedical applications of functionalised carbon nanotubes. *Chemical Communications*, *5*(5), 571–577. doi:10.1039/b410943k PMID:15672140

Birnbaum, L. S., Staskal, D. F., & Diliberto, J. J. (2003). Health effects of polybrominated dibenzo-p-dioxins (PBDDs) and dibenzofurans (PBDFs). *Environment International*, *29*(6), 855–860. doi:10.1016/S0160-4120(03)00106-5 PMID:12850101

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Bitencourt, M., & Freitas, M. P. (2008). MIA-QSAR evaluation of a series of sulfonylurea herbicides. *Pest Management Science*, *64*(8), 800–807. doi:10.1002/ps.1565 PMID:18338340

Blaney, J., & Hansch, C. (1990). Application of molecular graphics to the analysis of macromolecular structure. *Comprehensive Medicinal Chemistry*, *4*, 459–496.

Blankley, C. J. (1996). *Structure property correlations in drug research*. Academic Press.

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

Bohari, M. H., Srivastava, H. K., & Sastry, G. N. (2011). Analogue-based approaches in anti-cancer compound modelling: The relevance of QSAR models. *Organic and Medicinal Chemistry Letters*, *1*(1), 1–12. doi:10.1186/2191-2858-1-3 PMID:22373294

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

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

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

Bonaccorsi, R., Scrocco, E., & Tomasi, J. (1970). Molecular SCF calculations for the ground state of some three-membered ring molecules: $(CH_2)_3$, $(CH_2)_2NH$, $(CH_2)NH_2^+$, $(CH_2)_2O$, $(CH_2)_2S$, $(CH)_2CH_2$, and N_2CH_2. *The Journal of Chemical Physics*, *52*(10), 5270–5284. doi:10.1063/1.1672775

Bonchev, D. (1995). Topological order in molecules 1. Molecular branching revisited. *Journal of Molecular Structure THEOCHEM*, *336*(2-3), 137–156. doi:10.1016/0166-1280(94)04081-3

Bonchev, D., & Trinajsti, N. (1977). Information theory, distance matrix, and molecular branching. *The Journal of Chemical Physics*, *67*(10), 4517–4533. doi:10.1063/1.434593

Bondarenko, O., Juganson, K., Ivask, A., Kasemets, K., Mortimer, M., & Kahru, A. (2013). Toxicity of Ag, CuO and ZnO nanoparticles to selected environmentally relevant test organisms and mammalian cells in vitro: A critical review. *Archives of Toxicology*, *87*(7), 1181–1200. doi:10.1007/s00204-013-1079-4 PMID:23728526

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.

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.

Borges-Walmsley, M., McKeegan, K., & Walmsley, A. (2003). Structure and function of efflux pumps that confer resistance to drugs. *The Biochemical Journal*, *376*(2), 313–338. doi:10.1042/BJ20020957 PMID:13678421

Borm, P. J., Robbins, D., Haubold, S., Kuhlbusch, T., Fissan, H., & Donaldson, K. et al. (2006). The potential risks of nanomaterials: A review carried out for ECETOC. *Particle and Fibre Toxicology*, *3*(1), 11. doi:10.1186/1743-8977-3-11 PMID:16907977

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

Bréchignac, C., Houdy, P., & Lahmani, M. (Eds.). (2007). Nanomaterials and nanochemistry. Berlin: Springer Science+Bussiness Media. doi:10.1007/978-3-540-72993-8

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. 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., 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

Bro, R. (1996). Multiway calibration: Multilinear PLS. *Journal of Chemometrics*, *10*(1), 47–61. doi:10.1002/(SICI)1099-128X(199601)10:1<47::AID-CEM400>3.0.CO;2-C

Bro, R. (1997). PARAFAC: Tutorial and applications. *Chemometrics and Intelligent Laboratory Systems*, *38*(2), 149–171. doi:10.1016/S0169-7439(97)00032-4

Brown, R. D., & Martin, Y. C. (1997). The information content of 2D and 3D structural descriptors relevant to ligand-receptor binding. *Journal of Chemical Information and Computer Sciences*, *37*(1), 1–9. doi:10.1021/ci960373c

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

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

Brunner, T. J., Wick, P., Manser, P., Spohn, P., Grass, R. N., & Limbach, L. K. et al. (2006). In vitro cytotoxicity of oxide nanoparticles: Comparison to asbestos, silica, and the effect of particle solubility. *Environmental Science & Technology*, *40*(14), 4374–4381. doi:10.1021/es052069i PMID:16903273

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

Burello, E., & Worth, A. P. (2011). QSAR modeling of nanomaterials. *Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology*, *3*(3), 298–306. PMID:21384562

Bush, K., & Miller, G. H. (1998). Bacterial enzymatic resistance: β-lactamases and aminoglycoside-modifying enzymes. *Current Opinion in Microbiology*, *1*(5), 509–515. doi:10.1016/S1369-5274(98)80082-9 PMID:10066532

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

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

Buttingsrud, B., Alsberg, B. K., & Åstrand, P.-O. (2007a). An investigation of descriptors based on the critical points in the electron density by building quantitative structure–property relationships for proton chemical shifts. *Journal of Molecular Structure THEOCHEM*, *810*(1-3), 15–24. doi:10.1016/j.theochem.2007.01.031

Buttingsrud, B., Alsberg, B. K., & Åstrand, P.-O. (2007b). Quantitative prediction of the absorption maxima of azobenzene dyes from bond lengths and critical points in the electron density. *Physical Chemistry Chemical Physics*, *9*, 2226–2233. doi:10.1039/b617470a PMID:17487319

Buttingsrud, B., Alsberg, B. K., & Åstrand, P.-O. (2007c). Validation of critical points in the electron density as descriptors by building quantitative structure-property relationships for the atomic polar tensor. *Journal of Computational Chemistry*, *28*(13), 2130–2139. doi:10.1002/jcc.20666 PMID:17464968

Cai, R., Van, G. M., Aw, P. K., & Itoh, K. (2006). Solar-driven self-cleaning coating for a painted surface. *Comptes Rendus. Chimie*, *9*(5-6), 829–835. doi:10.1016/j.crci.2005.04.007

Camenisch, G., Folkers, G., & van de Waterbeemd, H. (1998). Shapes of membrane permeabilitylipophilicity curves: Extension of theoretical models with an aqueous pore pathway. *European Journal of Pharmaceutical Sciences*, *6*(4), 321–329. doi:10.1016/S0928-0987(98)00033-5

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

Cao, D. S., Xu, Q. S., Liang, Y. Z., Chen, X., & Li, H. D. (2010). Prediction of aqueous solubility of druglike organic compounds using partial least squares, back-propagation network and support vector machine. *Journal of Chemometrics*, *24*, 584–595.

Carlsen, L., Kenessov, B. N., & Batyrbekova, S. Y. (2009). A QSAR/QSTR study on the human health impact of the rocket fuel 1, 1-dimethyl hydrazine and its transformation products: Multicriteria hazard ranking based on partial order methodologies. *Environmental Toxicology and Pharmacology*, *27*(3), 415–423. doi:10.1016/j.etap.2009.01.005 PMID:21783973

Carlsson, C., Johansson, A. K., Alvan, G., Bergman, K., & Kuhler, T. (2006b). Are pharmaceuticals potent environmental pollutants? Part II: Environmental risk assessments of selected pharmaceutical excipients. *The Science of the Total Environment*, *364*(1-3), 88–95. doi:10.1016/j.scitotenv.2005.06.036 PMID:16260028

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

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

Castro, E. A., Tueros, M., & Toropov, A. A. (2000). Maximum topological distances based indices as molecular descriptors for QSPR 2 — Application to aromatic hydrocarbons. *Computers & Chemistry*, *24*(5), 571–576. doi:10.1016/S0097-8485(99)00095-9 PMID:10890366

Cattaneo, A. G., Gornati, R., Sabbioni, E., Chiriva-Internati, M., Cobos, E., Jenkins, M. R., & Bernardini, G. (2010). Nanotechnology and human health: Risks and benefits. *Journal of Applied Toxicology*, *30*(8), 730–744. doi:10.1002/jat.1609 PMID:21117037

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 nonsteroidal 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

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

Cayley, A. (1875). On the analytical forms called trees. *American Journal of Mathematics*, *4*(1), 266–268.

CEPA. (1999). *The Canadian environmental protection act the government of Canada*. Canada: CEPA.

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

Chakraborty, A., Duley, S., Giri, S., & Chattaraj, P. K. (2012). An understanding of the origin of chemical reactivity from a conceptual DFT approach. In J. Wiley (Ed.), A matter of density, exploring the electron density concept in the chemical, biological, and materials sciences. Academic Press. doi:10.1002/9781118431740.ch7

Chakraborty, A., Pan, S., & Chattaraj, P. K. (2013). Biological activity and toxicity: A conceptual DFT approach. *Structure and Bonding*, *150*, 143–179. doi:10.1007/978-3-642-32750-6_5

Chamberland, S., Ishida, Y., Lee, V. J., Leger, R., Nakayama, K., Ohta, T., et al. (2002). *US patent no. 6399629*. Washington, DC: US Patent Office.

Chamorro, E., Chattaraj, P. K., & Fuentealba, P. (2003). Variation of the electrophilicity index along the reaction path. *The Journal of Physical Chemistry A*, *107*(36), 7068–7072. doi:10.1021/jp035435y

Chang, C., Ray, A., & Swaan, P. (2005). *In silico* strategies for modeling membrane transporter function. *Drug Discovery Today*, *10*(9), 663–671. doi:10.1016/S1359-6446(05)03429-X PMID:15894231

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

Chattaraj, P. K. (1992). Electronegativity and hardness: A density functional treatment. *Journal of the Indian Chemical Society*, *69*, 173–183.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chattaraj, P. K. (1996). The maximum hardness principle: An overview. *Proceedings of the Indian National Science Academy - Part A*, *62*, 513–531.

Chattaraj, P. K. (Ed.). (2009). *Chemical reactivity theory: A density functional view*. Taylor and Francis/CRC Press. doi:10.1201/9781420065442

Chattaraj, P. K., Arun Murthy, T. V. S., Giri, S., & Roy, D. R. (2007). A connection between softness and magnetizability. *Journal of Molecular Structure THEOCHEM*, *813*(1-3), 63–65. doi:10.1016/j.theochem.2007.02.035

Chattaraj, P. K., Chakraborty, A., & Giri, S. (2009). Net electrophilicity. *The Journal of Physical Chemistry A*, *113*(37), 10068–10074. doi:10.1021/jp904674x PMID:19702288

Chattaraj, P. K., Fuentealba, P., Gomez, B., & Contreras, R. (2000). Woodward-Hoffmann rule in the light of the principles of maximum hardness and minimum polarizability: DFT and ab initio SCF studies. *Journal of the American Chemical Society*, *122*(2), 348–351. doi:10.1021/ja992337a

Chattaraj, P. K., Fuentealba, P., Jaque, P., & Toro-Labbé, A. (1999). Validity of the minimum polarizability principle in molecular vibrations and internal rotations: An ab initio SCF study. *The Journal of Physical Chemistry A*, *103*(46), 9307–9312. doi:10.1021/jp9918656

Chattaraj, P. K., Liu, G. H., & Parr, R. G. (1995). The maximum hardness principle in the Gyftopoulos-Hatsopoulos three-level model for an atomic or molecular species and its positive and negative ions. *Chemical Physics Letters*, *237*(1-2), 171–176. doi:10.1016/0009-2614(95)00280-H

Chattaraj, P. K., Maiti, B., & Sarkar, U. (2003). Philicity: A unified treatment of chemical reactivity and selectivity. *The Journal of Physical Chemistry A*, *107*(25), 4973–4975. doi:10.1021/jp034707u

Chattaraj, P. K., & Poddar, A. (1998). A density functional treatment of chemical reactivity and the associated electronic structure principles in the excited electronic states. *The Journal of Physical Chemistry A*, *102*(48), 9944–9948. doi:10.1021/jp982734s

Chattaraj, P. K., & Roy, D. R. (2007). Update 1 of: Electrophilicity index. *Chemical Reviews*, *107*(9), PR46–PR74. doi:10.1021/cr078014b PMID:21306180

Chattaraj, P. K., Roy, D. R., Geerlings, P., & Torrent-Sucarrat, M. (2007). Local hardness: A critical account. *Theoretical Chemistry Accounts*, *118*(5-6), 923–930. doi:10.1007/s00214-007-0373-8

Chattaraj, P. K., Roy, D. R., Giri, S., Mukherjee, S., Subramanian, V., & Parthasarathi, R. et al. (2007). An atom counting and electrophilicity based QSTR approach. *Journal of Chemical Sciences*, *119*(5), 475–488. doi:10.1007/s12039-007-0061-1

Chattaraj, P. K., Sarkar, U., & Roy, D. R. (2006). Electrophilicity index. *Chemical Reviews*, *106*(6), 2065–2091. doi:10.1021/cr040109f PMID:16771443

Chattaraj, P. K., & Sengupta, S. (1996). Popular electronic structure principles in a dynamical context. *Journal of Physical Chemistry*, *100*(40), 16126–16130. doi:10.1021/jp961096f

Chattaraj, P. K., & Sengupta, S. (1997). Dynamics of chemical reactivity indices for a many-electron system in its ground and excited states. *The Journal of Physical Chemistry A*, *101*(42), 7893–7900. doi:10.1021/jp971408u

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

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 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Chaudry, U. A., & Popelier, P. L. A. (2003). Ester hydrolysis rate constant prediction from quantum topological molecular similarity descriptors. *The Journal of Physical Chemistry A*, *107*(22), 4578–4582. doi:10.1021/jp034272a

Chaudry, U. A., & Popelier, P. L. A. (2004). Estimation of pKa using quantum topological molecular similarity descriptors: application to carboxylic acids, anilines and phenols. *The Journal of Organic Chemistry*, *69*(2), 233–241. doi:10.1021/jo0347415 PMID:14725434

Chau, Y. T., & Yap, C. W. (2012). Quantitative nanostructure–activity relationship modelling of nanoparticles. *RSC Advances*, *2*(22), 8489–8496. doi:10.1039/c2ra21489j

Chemical Computing Group. (n.d.). Retrieved from www.chemcomp.com

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

Chen, C. Y. (2011). TCM Database@Taiwan: The world's largest traditional Chinese medicine database for drug screening in silico. *PLoS ONE*, *6*(1), e15939. doi:10.1371/journal.pone.0015939 PMID:21253603

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

Chen, F. Q., & Gerion, D. (2004). Fluorescent CdSe/ZnS nanocrystal-peptide conjugates for long-term, nontoxic imaging and nuclear targeting in living cells. *Nano Letters*, *4*(10), 1827–1832. doi:10.1021/nl049170q

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

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

Chen, J. J., Tsai, C. A., Young, J. F., & Kodell, R. L. (2005). Classification ensembles for unbalanced class sizes in predictive toxicology. *SAR and QSAR in Environmental Research*, *16*(6), 517–529. doi:10.1080/10659360500468468 PMID:16428129

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

Chen, J., Morita, Y., Huda, M. N., Kuroda, T., Mizushima, T., & Tsuchiya, T. (2002). VmrA, a member of a novel class of Na^+-coupled multidrug efflux pumps from *Vibrio parahaemolyticus*. *Journal of Bacteriology*, *184*(2), 572–576. doi:10.1128/JB.184.2.572-576.2002 PMID:11751837

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

Chen, S. (1998). Aromatase and breast cancer. *Frontiers in Bioscience*, *3*, 922–933. PMID:9696881

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, 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

Chen, X.-Q., Cho, S. J., Li, Y., & Venkatesh, S. (2002). Prediction of aqueous solubility of organic compounds using a quantitative structure–property relationship. *Journal of Pharmaceutical Sciences*, *91*(8), 1838–1852. doi:10.1002/jps.10178 PMID:12115811

Chen, Y., Chen, J., Dong, J., & Jin, Y. (2004). Comparing study of the effect of nanosized silicon dioxide and microsized silicon dioxide on fibrogenesis in rats. *Toxicology and Industrial Health*, *20*(1), 21–27. doi:10.1191/0748233704th190oa PMID:15807405

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

Chevalier, J., Bredin, J., Mahamoud, A., Malléa, M., Barbe, J., & Pagès, J.-M. (2004). Inhibitors of antibiotic efflux in resistant Enterobacter aerogenes and Klebsiella pneumoniae strains. *Antimicrobial Agents and Chemotherapy*, *48*(3), 1043–1046. doi:10.1128/AAC.48.3.1043-1046.2004 PMID:14982806

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

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

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

Chung, Y., & Saier, M. Jr. (2001). SMR-type multidrug resistance pumps. *Current Opinion in Drug Discovery & Development*, *4*(2), 237–245. PMID:11378963

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

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., & 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

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

Clark, K. A., White, R. H., & Silbergeld, E. K. (2011). Predictive models for nanotoxicology: Current challenges and future opportunities. *Regulatory Toxicology and Pharmacology*, *59*(3), 361–363. doi:10.1016/j.yrtph.2011.02.002 PMID:21310205

Coban, A. Y., Ekinci, B., & Durupinar, B. (2004). A multidrug efflux pump inhibitor reduces fluoroquinolone resistance in *Pseudomonas aeruginosa* isolates. *Chemotherapy*, *50*(1), 22–26. doi:10.1159/000077280 PMID:15084801

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Coi, A., Massarelli, I., Murgia, L., Saraceno, M., Calderone, V., & Bianuccia, A. M. (2006). Prediction of hERG potassium channel affinity by the CODESSA approach. *Bioorganic & Medicinal Chemistry*, *14*(9), 3153–3159. doi:10.1016/j.bmc.2005.12.030 PMID:16426850

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

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

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

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

Concu, R., Dea-Ayuela, M. A., Perez-Montoto, L. G., Prado-Prado, F. J., & Uriarte, E. (2009). 3D entropy and moments prediction of enzyme classes and experimental-theoretic study of peptide fingerprints in Leishmania parasites. *Biochimica et Biophysica Acta*, (12), 1784–1794. PMID:19716935

Concu, R., Podda, G., Uriarte, E., & Gonzalez-Diaz, H. (2009). Computational chemistry study of 3D-structure-function relationships for enzymes based on Markov models for protein electrostatic, HINT, and van der Waals potentials. *Journal of Computational Chemistry*, *30*(9), 1510–1520. doi:10.1002/jcc.21170 PMID:19086060

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

Consonni, V., Todeschini, R., & Pavan, M. (2002). Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. Part 1. Theory of the novel 3D molecular descriptors. *Journal of Chemical Information and Computer Sciences*, *42*(3), 682–692. doi:10.1021/ci015504a PMID:12086530

Cooper, E. R., Siewicki, T. C., & Phillips, K. (2008). Preliminary risk assessment database and risk ranking of pharmaceuticals in the environment. *The Science of the Total Environment*, *398*(1-3), 26–33. doi:10.1016/j.scitotenv.2008.02.061 PMID:18440598

Cormanich, R. A., Goodarzi, M., & Freitas, M. P. (2009). Improvement of multivariate image analysis applied to quantitative structure–activity relationship (QSAR) analysis by using wavelet-principal component analysis ranking variable selection and least-squares support vector machine regression: QSAR study of checkpoint kinase WEE1 inhibitors. *Chemical Biology & Drug Design*, *73*(2), 244–252. doi:10.1111/j.1747-0285.2008.00764.x PMID:19207427

Cormanich, R., Nunes, C. A., & Freitas, M. P. (2012). Chemical drawings correlate to biological properties: MIA-QSAR. *Quimica Nova*, *35*, 1157–1163. doi:10.1590/S0100-40422012000600017

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. doi:10.1007/BF00994018

Coulson, C. A. (1939). The electronic structure of some polyenes and aromatic molecules. VII. Bonds of fractional order by the molecular orbital method. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, *16*(938), 413–428. doi:10.1098/rspa.1939.0006

Counsell, R. E., Klimstra, P. D., & Colton, F. B. (1962). Anabolic agents. Derivatives of 5α-Androst-1-ene. *The Journal of Organic Chemistry*, *27*(1), 248–253. doi:10.1021/jo01048a060

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.

Cox, D. S., Gao, H., Raje, S., Scott, K. R., & Eddington, N. D. (2001a). Enhancing the permeation of marker compounds and enaminone anticonvulsants across Caco-2 monolayers by modulating tight junctions using zonula occludens toxin. *European Journal of Pharmaceutics and Biopharmaceutics*, *52*(2), 145–150. doi:10.1016/S0939-6411(01)00172-2 PMID:11522479

Cox, D. S., Scott, K. R., Gao, H., Raje, S., & Eddington, N. D. (2001b). Influence of multidrug resistance (MDR) proteins at the blood-brain barrier on the transport and brain distribution of enaminone anticonvulsants. *Journal of Pharmaceutical Sciences*, *90*(10), 1540–1552. doi:10.1002/jps.1104 PMID:11745712

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. III, & Milne, M. (1979). Abstracts of papers of the American Chemical Society. *Computer Chemistry Section, 44*.

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

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

Cronin, M. T. D. (2010). Quantitative structure-activity relationships (QSAR) - Applications and methodology. In T. Puzyn, J. Leszczynski, & M. T. D. Cronin (Eds.), *Recent advances in QSAR studies. Challenges and advances in computational chemistry and physics*. Dordrecht, The Netherlands: Springer. doi:10.1007/978-1-4020-9783-6_1

Cronin, M. T. D., Aptula, A. O., Judith, C., Duffy, J. C., Netzeva, T. I., & Rowe, P. H. et al. (2002). Comparative assessment of methods to develop QSARs for the prediction of the toxicity of phenols to Tetrahymena pyriformis. *Chemosphere*, *49*(10), 1201–1221. doi:10.1016/S0045-6535(02)00508-8 PMID:12489717

Cronin, M. T. D., Gregory, B. W., & Schultz, T. W. (1998). Quantitative structure-activity analyses of nitrobenzene toxicity to tetrahymena pyriformis. *Chemical Research in Toxicology*, *11*(8), 902–908. doi:10.1021/tx970166m PMID:9705752

Cronin, M. T. D., & Schultz, T. W. (1996). Structure-toxicity relationships for phenols to *Tetrahymena pyriformis*. *Chemosphere*, *32*(8), 1453–1468. doi:10.1016/0045-6535(96)00054-9 PMID:8653384

Cronin, M. T. D., & Schultz, T. W. (1998). Structure–toxicity relationships for three mechanisms of action of toxicity to *vibrio fischeri*. *Ecotoxicology and Environmental Safety*, *39*(1), 65–69. doi:10.1006/eesa.1997.1618 PMID:9515077

Cronin, M. T. D., & Schultz, T. W. (2003). Pitfalls in QSAR. *Journal of Molecular Structure THEOCHEM*, *622*(1-2), 39–51. doi:10.1016/S0166-1280(02)00616-4

Cros, A. F. A. (1863, January 9). *Action de l'alcohol amylique sur l'organisme*. Faculty of Medicine, University of Strasbourg.

Crosland, M. P. (1959). The use of diagrams as chemical 'equations' in the lecture notes of William Cullen and Joseph Black. *Annals of Science*, *15*(2), 75–90. doi:10.1080/00033795900200088

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

Cruciani, G., Pastor, M., & Guba, W. (2000). VolSurf: A new tool for the pharmacokinetic optimization of lead compounds. *European Journal of Pharmaceutical Sciences*, *11*, S29–S39. doi:10.1016/S0928-0987(00)00162-7 PMID:11033425

Crum-Brown, A., & Fraser, T. R. (1869). On the connection between chemical constitution and physiological action. Part 1: On the physiological action of the ammonium bases derived from Strychia, Brucia, Thebia, Codeia, Morphia, and Nicotia. *Transactions of the Royal Society of Edinburgh*, *25*, 257–274.

Cruz-Monteagudo, M., Ancede-Gallardo, E., Jorge, M., & Cordeiro, M. N. D. S. (2013). Chemoinformatics profiling of ionic liquids--automatic and chemically interpretable cytotoxicity profiling, virtual screening, and cytotoxicophore identification. *Toxicological Sciences*, *136*(2), 548–565. doi:10.1093/toxsci/kft209 PMID:24068674

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

Cruz-Monteagudo, M., Gonzalez-Diaz, H., Borges, F., Dominguez, E. R., & Cordeiro, M. N. D. S. (2008). 3D-MEDNEs: An alternative "in silico" technique for chemical research in toxicology. 2. quantitative proteome-toxicity relationships (QPTR) based on mass spectrum spiral entropy. *Chemical Research in Toxicology*, *21*(3), 619–632. doi:10.1021/tx700296t PMID:18257557

Cruz-Monteagudo, M., Gonzalez-Diaz, H., Borges, F., & Gonzalez-Diaz, Y. (2006). Simple stochastic fingerprints towards mathematical modeling in biology and medicine. 3. Ocular irritability classification model. *Bulletin of Mathematical Biology*, *68*(7), 1555–1572. doi:10.1007/s11538-006-9083-y PMID:16865609

CSTEE. (2001). Environmental risk assesment of medicinal products for human use. Brussels, Belgium: Author.

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

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

da Silva Junkes, B., Arruda, A. C. S., Yunes, R. A., Porto, L. C., & Heinzen, V. E. F. (2005). Semi-empirical topological index: A tool for QSPR/QSAR studies. *Journal of Molecular Modeling*, *11*(2), 128–134. doi:10.1007/s00894-004-0231-8 PMID:15723207

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

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

Dai, Y., Zhang, X., Zhang, X., Wang, H., & Lu, Z. (2008). DFT and GA studies on the QSAR of 2-aryl-5-nitro-1H-indole derivatives as NorA efflux pump inhibitors. *Journal of Molecular Modeling*, *14*(9), 807–812. doi:10.1007/s00894-008-0328-6 PMID:18575902

Dalgarno, A. (1962). Atomic polarizabilities and shielding factors. *Advances in Physics*, *11*(44), 281–315. doi:10.1080/00018736200101302

Dalton, J. (1808). New system of chemical philosophy. London: Chembridge Press.

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:10.2174/1389557511313666010 4 PMID:24195665

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Damme, S. V., & Bultinck, P. (2010). 3D QSAR based on conceptual DFT molecular fields: Antituberculotic activity. *Journal of Molecular Structure THEOCHEM*, *943*(1-3), 83–89. doi:10.1016/j.theochem.2009.10.031

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

Das, R. N., & Roy, K. (2012). Development of classification and regression models for *Vibrio fischeri* toxicity of ionic liquids: Green solvents for the future. *Toxicological Reviews*, *1*, 186–195.

Das, R. N., & Roy, K. (2013). QSPR with extended topochemical atom (ETA) indices. 4. Modeling aqueous solubility of drug like molecules and agrochemicals following OECD guidelines. *Structural Chemistry*, *24*(1), 303–331. doi:10.1007/s11224-012-0080-5

Das, R. N., & Roy, K. (2014a). Predictive modeling studies for the ecotoxicity of ionic liquids towards the green algae *Scenedesmus vacuolatus. Chemosphere*, *104*, 170–176. doi:10.1016/j.chemosphere.2013.11.002 PMID:24296027

Das, R. N., & Roy, K. (2014b). Predictive *in silico* modeling of ionic liquids toward inhibition of the acetyl cholinesterase enzyme of *electrophorus electricus*: A predictive toxicology approach. *Industrial & Engineering Chemistry Research*, *53*(2), 1020–1032. doi:10.1021/ie403636q

Das, R. N., Sanderson, H., Mwambo, A. E., & Roy, K. (2013). Preliminary studies on model development for rodent toxicity and its interspecies correlation with aquatic toxicities of pharmaceuticals. *Bulletin of Environmental Contamination and Toxicology*, *90*(3), 375–381. doi:10.1007/s00128-012-0921-3 PMID:23238824

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

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

Daughton, C. G., & Ruhoy, I. S. (2009). Environmental footprint of pharmaceuticals: The significance of factors beyond direct excretion to sewers. *Environmental Toxicology and Chemistry*, *28*(12), 2495–2521. doi:10.1897/08-382.1 PMID:19382823

Daughton, C. G., & Ruhoy, I. S. (2011). Green pharmacy and pharmEcovigilance: Prescribing and the planet. *Expert Review and Clinical Pharmacology*, *4*(2), 211–232. doi:10.1586/ecp.11.6 PMID:22115404

Davis, L. (Ed.). (1991). *Handbook of genetic algorithms*. London: Van Nostrand-Reinhold.

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

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

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

de Melo, E. B. (2012). A new quantitative structure–property relationship model to predict bioconcentration factors of polychlorinated biphenyls (PCBs) in fishes using E-state index and topological descriptors. *Ecotoxicology and Environmental Safety*, *75*, 213–222. doi:10.1016/j.ecoenv.2011.08.026 PMID:21959189

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

de Melo, E. B., & Ferreira, M. M. C. (2009). Multivariate QSAR study of 4,5-dihydroxypyrimidine carboxamides as HIV-1 integrase inhibitors European. *Journal of Medicinal Chemistry*, *44*(9), 3577–3583. doi:10.1016/j.ejmech.2009.03.001 PMID:19327872

De Volder, M. F., Tawfick, S. H., Baughman, R. H., & Hart, A. J. (2013). Carbon nanotubes: Present and future commercial applications. *Science*, *339*(6119), 535–539. doi:10.1126/science.1222453 PMID:23372006

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

Debnath, A. K., Shusterman, A. J., Lopez de Compadre, R. L., & Hasch, C. (1994). The importance of the hydrophobic interaction in the mutagenicity of organic compounds. *Mutation Research/Fundamental and Molecular Mechanisms, 305*, 63-72.

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

Deeb, O., Alfalah, S., Freitas, M. P., da Cunha, E. F. F., & Ramalho, T. C. (2012). Exploring MIA-QSARs for farnesyltransferase inhibitory effect of antimalarial compounds refined by docking simulations. *Journal of Biophysical Chemsitry*, *3*(01), 58–71. doi:10.4236/jbpc.2012.31008

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

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

Delay, M., & Frimmel, F. H. (2012). Nanoparticles in aquatic systems. *Analytical and Bioanalytical Chemistry*, *402*(2), 583–592. doi:10.1007/s00216-011-5443-z PMID:22038580

Delmar, J. A., Su, C.-C., & Yu, E. W. (2014). Bacterial multidrug efflux transporters. *Annual Review of Biophysics, 43*, 93-117. doi: 10.1146/annurev-biophys-051013-022855

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

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

Dennington, R. D., II, Keith, T. A., & Millam, J. M. (2008). GaussView 5.0. Wallingford.

DePristo, M. A., de Bakker, P. I., & Blundell, T. L. (2004). Heterogeneity and inaccuracy in protein structures solved by X-ray crystallography. *Structure (London, England)*, *12*(5), 831–838. doi:10.1016/j.str.2004.02.031 PMID:15130475

Devillers, J., & Balaban, A. T. (1999). *Topological indices and related descriptors in QSAR and QSPR*. Amsterdam: Gordon Breach Scientific Publishers.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Devillers, J., & Chambon, P. (1986). Acute toxicity and QSAR of chlorophenols on *Daphnia magna. Bulletin of Environmental Contamination and Toxicology*, *37*(1), 599–605. doi:10.1007/BF01607810 PMID:3768559

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.

Dimitrov, S., Dimitrova, G., Pavlov, T., Dimitrova, N., Patlewicz, G., Niemela, J., & Mekenyan, O. A. (2005). Stepwise approach for defining the applicability domain of SAR and QSAR models. *Journal of Chemical Information and Modeling*, *45*(4), 839–849. doi:10.1021/ci0500381 PMID:16045276

Diudea, M. V., & Gutman, I. (1998). Wiener-type topological indices. *Croatica Chemica Acta*, *71*(1), 21–51.

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

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

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

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

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

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

Dragon Version 6 TALETE srl, Italy. (2010). Retrieved April 30, 2014, http://www.talete.mi.it/products/dragon_description.htm

Dragos, H., Gilles, M., & Alexandre, V. (2009). Predicting the predictability: A unified approach to the applicability domain problem of QSAR models. *Journal of Chemical Information and Modeling*, *49*(7), 1762–1776. doi:10.1021/ci9000579 PMID:19530661

Dreher, K. L. (2004). Health and environmental impact of nanotechnology: Toxicological assessment of manufactured nanoparticles. *Toxicological Sciences*, *77*(1), 3–5. doi:10.1093/toxsci/kfh041 PMID:14756123

Drexler, K. E. (1992). Nanotechnology: The past and the future. *Science*, *255*(5042), 268–269. doi:10.1126/science.255.5042.268-b PMID:17779552

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

Du, B., Zheng, M., Huang, Y., Liu, A., Tian, H., Li, L., & Su, G. et al. (2010). Mixed polybrominated/chlorinated dibenzo-p-dioxins and dibenzofurans in stack gas emissions from industrial thermal processes. *Environmental Science & Technology*, *44*(15), 5818–5823. doi:10.1021/es100867d PMID:20586446

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Duchowicz, P. R., & Castro, E. A. (2009). QSPR studies on aqueous solubilities of drug-like compounds. *International Journal of Molecular Sciences*, *10*(6), 2558–2577. doi:10.3390/ijms10062558 PMID:19582218

Duchowicz, P. R., Castro, E. A., Fernandez, F. M., & Gonzalez, M. P. (2005). A new search algorithm for QSPR/QSAR theories: Normal boiling points of some organic molecules. *Chemical Physics Letters*, *412*(4-6), 376–380. doi:10.1016/j.cplett.2005.07.016

Duchowicz, P. R., Comelli, N. C., Ortiz, E. V., & Castro, E. A. (2012). QSAR study for carcinogenicity in a large set of organic compounds. *Current Drug Safety*, *7*, 282–288. doi:10.2174/157488612804096623 PMID:23062240

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

Du-Cuny, L., Huwyler, J., Wiese, M., & Kansy, M. (2008). Computational aqueous solubility prediction for drug-like compounds in congeneric series. *European Journal of Medicinal Chemistry*, *43*(3), 501–512. doi:10.1016/j.ejmech.2007.04.009 PMID:17574307

Du, D., Wang, Z., James, N. R., Voss, J. E., Klimont, E., & Ohene-Agyei, T. et al. (2014). Structure of the AcrAB-TolC multidrug efflux pump. *Nature*, *509*(7501), 512–515. doi:10.1038/nature13205 PMID:24747401

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

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

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

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

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

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

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

Durdagi, S., Mavromoustakos, T., & Papadopoulos, M. G. (2008). 3D QSAR CoMFA/CoMSIA, molecular docking and molecular dynamics studies of fullerene-based HIV-1 PR inhibitors. *Bioorganic & Medicinal Chemistry Letters*, *18*(23), 6283–6289. doi:10.1016/j.bmcl.2008.09.107 PMID:18951793

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

Du, X., Guo, C., Hansell, E., Doyle, P. S., Caffrey, C. R., & Holler, T. P. et al. (2002). Synthesis and structure-activity relationship study of potent trypanocidal thio semicarbazone inhibitors of the trypanosomal cysteine protease cruzain. *Journal of Medicinal Chemistry*, *45*(13), 2695–2707. doi:10.1021/jm010459j PMID:12061873

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

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

EC. (2006). *Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerming the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, and amemding Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC*. Brussels: The European Parliament and the Council of the European Union.

Edafiogho, I. O., Ananthalakshmi, K. V., & Kombian, S. B. (2006). Anticonvulsant evaluation and mechanism of action of benzylamino enaminones. *Bioorganic & Medicinal Chemistry*, *14*(15), 5266–5272. doi:10.1016/j.bmc.2006.03.049 PMID:16621576

Eddington, N. D., Cox, D. S., Roberts, R. R., Butcher, R. J., Edafiogho, I. O., & Stables, J. P. et al. (2002). Synthesis and anticonvulsant activity of enaminones. 4. Investigations on isoxazole derivatives. *European Journal of Medicinal Chemistry*, *37*(8), 635–648. doi:10.1016/S0223-5234(02)01377-6 PMID:12161061

Edward, W. Y., Aires, J. R., McDermott, G., & Nikaido, H. (2005). A periplasmic drug-binding site of the AcrB multidrug efflux pump: A crystallographic and site-directed mutagenesis study. *Journal of Bacteriology*, *187*(19), 6804–6815. doi:10.1128/JB.187.19.6804-6815.2005 PMID:16166543

Edward, W. Y., McDermott, G., Zgurskaya, H. I., Nikaido, H., & Koshland, D. E. (2003). Structural basis of multiple drug-binding capacity of the AcrB multidrug efflux pump. *Science*, *300*(5621), 976–980. doi:10.1126/science.1083137 PMID:12738864

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

Eisenbrand, G., Pool-Zobel, B., Baker, V., Balls, M., Blaauboer, B. J., & Boobis, A. et al. (2002). Methods of in vitro toxicology. *Food and Chemical Toxicology*, *40*(2-3), 193–236. doi:10.1016/S0278-6915(01)00118-1 PMID:11893398

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

Elkins, C. A., & Nikaido, H. (2003). 3D structure of AcrB: The archetypal multidrug efflux transporter of *Escherichia coli* likely captures substrates from periplasm. *Drug Resistance Updates*, *6*(1), 9–13. doi:10.1016/S1368-7646(03)00004-9 PMID:12654283

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

Enalos InSilicoNano Platform. (n.d.). Retrieved from www.insilicotox.com

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

Epa, V. C., Burden, F. R., Tassa, C., Weissleder, R., Shaw, S., & Winkler, D. A. (2012). Modeling biological activities of nanoparticles. *Nano Letters*, *12*(11), 5808–5812. doi:10.1021/nl303144k PMID:23039907

EPISuite. (2010). Estimation programs interface suite™ for Microsoft® Windows, v 4.00. United States Environmental Protection Agency.

Eriksson, L., Jaworska, J., Worth, A. P., Cronin, M. T., 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

Eriksson, L., Johansson, E., Kettaneh-Wold, N., & Wold, S. (2001). *Multi- and megavariate data analysis. principles and applications*. Umea: Umetrics.

Eroglu, E., Palaz, S., Oltulu, O., Turkmen, H., & Ozaydın, C. (2007). Comparative QSTR study using semi-empirical and first principle methods based descriptors for acute toxicity of diverse organic compounds to the fathead minnow. *International Journal of Molecular Sciences*, *8*(12), 1265–1283. doi:10.3390/ijms8121265

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

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

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Estrada, E., Molina, E., & Perdomo-López, I. (2001). Can 3D structural parameters be predicted from 2D (topological) molecular descriptors? *Journal of Chemical Information and Computer Sciences*, *41*(4), 1015–1021. doi:10.1021/ci000170v PMID:11500118

Estrada, E., Molina, E., & Uriarte, E. (2001). Quantitative structure-toxicity relationships using TOPS-MODE. 2. Neurotoxicity of a non-congeneric series of solvents. *SAR and QSAR in Environmental Research*, *12*(5), 445–459. doi:10.1080/10629360108035384 PMID:11813810

Estrada, E., & Peña, A. (2000). In silico studies for the rational discovery of anticonvulsant compounds. *Bioorganic & Medicinal Chemistry*, *8*(12), 2755–2770. doi:10.1016/S0968-0896(00)00204-2 PMID:11131167

Estrada, E., Pena, A., & Garcia-Domenech, R. (1998). Designing sedative/hypnotic compounds from a novel substructural graph-theoretical approach. *Journal of Computer-Aided Molecular Design*, *12*(6), 583–595. doi:10.1023/A:1008048003720 PMID:9879506

Estrada, E., Quincoces, J. A., & Patlewicz, G. (2004). Creating molecular diversity from antioxidants in Brazilian propolis. Combination of TOPS-MODE QSAR and virtual structure generation. *Molecular Diversity*, *8*(1), 21–33. doi:10.1023/B:MODI.0000006804.97390.40 PMID:14964785

Estrada, E., & Uriarte, E. (2001). Quantitative structure-toxicity relationships using TOPS-MODE. 1. Nitrobenzene toxicity to Tetrahymena pyriformis. *SAR and QSAR in Environmental Research*, *12*(3), 309–324. doi:10.1080/10629360108032919 PMID:11696927

Estrada, E., Uriarte, E., Molina, E., Simon-Manso, Y., & Milne, G. W. (2006). An integrated in silico analysis of drug-binding to human serum albumin. *Journal of Chemical Information and Modeling*, *46*(6), 2709–2724. doi:10.1021/ci600274f PMID:17125211

Estrada, E., Uriarte, E., Montero, A., Teijeira, M., Santana, L., & De Clercq, E. (2000). A novel approach for the virtual screening and rational design of anticancer compounds. *Journal of Medicinal Chemistry*, *43*(10), 1975–1985. doi:10.1021/jm991172d PMID:10821710

Euler, L. (1736). Solutio problematis ad geometriam situs pertinentis. *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, *8*, 128–140.

Falandysz, J. (1999). *Polichlorowane bifenyle (PCBs) w środowisku. chemia, analiza, toksyczność, stężenia i ocena ryzyka*. Gdańsk: Fundacja Rozwoju Uniwersytetu Gdańskiego.

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

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

Farkas, O., Jakus, J., & Héberger, K. (2004). Quantitative structure–antioxidant activity relationships of flavonoid compounds. *Molecules (Basel, Switzerland)*, *9*(12), 1079–1088. doi:10.3390/91201079 PMID:18007505

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

Fechner, N., Jahn, A., Hinselmann, G., & Zell, A. (2009). Atomic local neighborhood flexibility incorporation into a structured similarity measure for QSAR. *Journal of Chemical Information and Modeling*, *49*(3), 549–560. doi:10.1021/ci800329r PMID:19434895

Fechner, N., Jahn, A., Hinselmann, G., & Zell, A. (2010). Estimation of the applicability domain of kernel-based machine learning models for virtual screening. *Journal of Cheminformatics*, *2*(1), 2. doi:10.1186/1758-2946-2-2 PMID:20222949

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Fenner, K., Scheringer, M., Macleod, M., Matthies, M., McKone, T., Stroebe, M., & Wania, F. et al. (2005). Comparing estimates of persistence and long-range transport potential among multimedia models. *Environmental Science & Technology*, *39*(7), 1932–1942. doi:10.1021/es048917b PMID:15871221

Ferguson, A. M., Heritage, T. W., Jonathon, P., Pack, S. E., Phillips, L., Rogan, J., & Snaith, P. J. (1997). EVA: A new theoretically based molecular descriptor for use in QSAR\QSPR analysis. *Journal of Computer-Aided Molecular Design*, *11*(2), 143–152. doi:10.1023/A:1008026308790 PMID:9089432

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), 387–403. doi:10.1098/rspb.1939.0030

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

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

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

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

Ferreira, M. M. C. (2001). Polycyclic aromatic hydrocarbons: A QSPR study. *Chemosphere*, *44*(2), 125–146. doi:10.1016/S0045-6535(00)00275-7 PMID:11444294

Ferreira, M., & Kiralj, R. (2004). QSAR study of β-lactam antibiotic efflux by the bacterial multidrug resistance pump AcrB. *Journal of Chemometrics*, *18*(5), 242–252. doi:10.1002/cem.867

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

Feynman, R. P. (1960). There's plenty of room at the bottom (an invitation to enter a new field of physics). *Engineering and Science*, *23*(5), 22–36.

Filho, E. B. A., Moraes, I. A., Weber, K. C., Rocha, G. B., & Vasconcellos, M. L. A. A. (2012). DFT/PCM, QTAIM, 1H NMR conformational studies and QSAR modeling of thirty-two anti-Leishmania amazonensis Morita–Baylis–Hillman Adducts. *Journal of Molecular Structure THEOCHEM*, *1022*, 72–80. doi:10.1016/j.molstruc.2012.04.051

Fischer, N. W. (1974). Kekulé and organic classification. *Ambix*, *21*(1), 29–52. doi:10.1179/amb.1974.21.1.29

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

Forina, M., Armanino, C., Leardi, R., & Drava, G. (1991). A class-modelling technique based on potential functions. *Journal of Chemometrics*, *5*(5), 435–453. doi:10.1002/cem.1180050504

Fouchécourt, M.-O., Béliveau, M., & Krishnan, K. (2001). Quantitative structure-pharmacokinetic relationship modeling. *The Science of the Total Environment*, *274*(1-3), 125–135. doi:10.1016/S0048-9697(01)00743-4 PMID:11453289

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Fourches, D., Pu, D., Tassa, C., Weissleder, R., Shaw, S. Y., Mumper, R. J., & Tropsha, A. (2010). Quantitative nanostructure-activity relationship modeling. *ACS Nano*, *4*(10), 5703–5712. doi:10.1021/nn1013484 PMID:20857979

Fourches, D., Pu, D., & Tropsha, A. (2011). Exploring quantitative nanostructure-activity relationships (QNAR) modeling as a tool for predicting biological effects of manufactured nanoparticles. *Combinatorial Chemistry & High Throughput Screening*, *14*(3), 217–225. doi:10.2174/138620711794728743 PMID:21275889

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

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

Freitas, M. P. (2006). MIA-QSAR modelling of anti-HIV-1 activities of some 2-amino-6-arylsulfonylbenzonitriles and their thio and sulfinyl congeners. *Organic & Biomolecular Chemistry*, *4*(6), 1154–1159. doi:10.1039/b516396j PMID:16525561

Freitas, M. P., Brown, S. D., & Martins, J. A. (2005). MIA-QSAR: A simple 2D image-based approach for quantitative structure-activity relationship analysis. *Journal of Molecular Structure*, *738*(1-3), 149–154. doi:10.1016/j.molstruc.2004.11.065

Freitas, M. P., da Cunha, E. F. F., Ramalho, T. C., & Goodarzi, M. (2008). Multimode methods applied on MIA descriptors in QSAR. *Current Computer-aided Drug Design*, *4*(4), 273–282. doi:10.2174/157340908786786038

Freitas, M. P., & Ramalho, T. C. (2013). Employing conformational analysis in the molecular modeling of agrochemicals: Insights on QSAR parameters of 2,4-D. *Ciência & Agrotecnologia*, *37*(6), 485–494. doi:10.1590/S1413-70542013000600001

Freitas, M. P., & Rittner, R. (2008). MIA-QSAR as an alternative approach for modeling some antifungals. *QSAR & Combinatorial Science*, *27*(5), 582–585. doi:10.1002/qsar.200710047

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

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.

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

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

Fuentealba, P., & Reyes, O. (1993). Atomic softness and the electric dipole polarizability. *Journal of Molecular Structure THEOCHEM*, *282*(1-2), 65–70. doi:10.1016/0166-1280(93)85035-W

Fujita, T. (1990). The extrathermodynamic approach to drug design. In Comprehensive medicinal chemistry (Vol. 4). Pergamon.

Fujita, T., & Ban, T. (1971). Structure-activity relation. 3. 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

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

Fukui, K. (1975). *Theory of orientation and stereoselection*. Berlin: Springer. doi:10.1007/978-3-642-61917-5

Fukui, K. (1982). Role of frontier orbitals in chemical reactions. *Science*, *218*(4574), 747–754. doi:10.1126/science.218.4574.747 PMID:17771019

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Fukui, K., Yonezawa, T., Nagata, C., & Shingu, H. (1954). Molecular orbital theory of orientation in aromatic, heteroaromatic, and other conjugated molecules. *The Journal of Chemical Physics*, *22*(8), 1433–1442. doi:10.1063/1.1740412

Fukui, K., Yonezawa, T., & Shingu, H. (1952). A molecular orbital theory of reactivity in aromatic hydrocarbons. *The Journal of Chemical Physics*, *20*(4), 722–725. doi:10.1063/1.1700523

Fukui, K., Yonezawa, Y., & Shingu, H. (1954). Theory of substitution in conjugated molecules. *Bulletin of the Chemical Society of Japan*, *27*(7), 423–427. doi:10.1246/bcsj.27.423

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

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

Gajewicz, A., Haranczyk, M., & Puzyn, T. (2010). Predicting logarithmic values of the subcooled liquid vapor pressure of halogenated persistent organic pollutants with QSPR. How different are chlorinated and brominated congeners? *Atmospheric Environment*, *44*(11), 1428–1436. doi:10.1016/j.atmosenv.2010.01.041

Gajewicz, A., Rasulev, B., Dinadayalane, T. C., Urbaszek, P., Puzyn, T., Leszczynska, D., & Leszczynsk, J. (2012). Advancing risk assessment of engineered nanomaterials: Application of computational approaches. *Advanced Drug Delivery Reviews*, *64*(15), 1663–1693. doi:10.1016/j.addr.2012.05.014 PMID:22664229

Gálvez, J. (1998). On a topological interpretation of electronic and vibrational molecular energies. *Journal of Molecular Structure THEOCHEM*, *429*, 255–264. doi:10.1016/S0166-1280(97)00366-7

Gao, H., Katzenellenbogen, J. A., Garg, R., & Hansch, C. (1999). Comparative QSAR analysis of estrogen receptor ligands. *Chemical Reviews*, *99*(3), 723–744. doi:10.1021/cr980018g PMID:11749430

García, A., Espinosa, R., Delgado, L., Casals, E., González, E., & Puntes, V. et al. (2011). Acute toxicity of cerium oxide, titanium oxide and iron oxide nanoparticles using standardized tests. *Desalination*, *269*(1-3), 136–141. doi:10.1016/j.desal.2010.10.052

Garcia, I., Fall, Y., Gomez, G., & Gonzalez-Diaz, H. (2011). First computational chemistry multi-target model for anti-Alzheimer, anti-parasitic, anti-fungi, and anti-bacterial activity of GSK-3 inhibitors in vitro, in vivo, and in different cellular lines. *Molecular Diversity*, *15*(2), 561–567. doi:10.1007/s11030-010-9280-3 PMID:20931280

García, J., Duchowicz, P. R., Rozas, M. F., Caram, J. A., Mirífico, M. V., Fernández, F. M., & Castro, E. A. (2011). A comparative QSAR on 1,2,5-thiadiazolidin-3-one 1,1-dioxide compounds as selective inhibitors of human serine proteinases. *Journal of Molecular Graphics & Modelling*, *31*, 10–19. doi:10.1016/j.jmgm.2011.07.007 PMID:21908217

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

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

Garg, R., Gupta, S. P., Gao, H., Babu, M. S., Debnath, A. K., & Hansch, C. (1999). Comparative quantitative structure-activity relationship studies on anti-HIV drugs. *Chemical Reviews*, *99*(12), 3525–3602. doi:10.1021/cr9703358 PMID:11849030

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Garkani-Nejad, Z., & Ahmadi-Roudi, B. (2010). Modeling the antileishmanial activity screening of 5-nitro-2-heterocyclic benzylidene hydrazides using different chemometrics methods. *European Journal of Medicinal Chemistry*, *45*(2), 719–726. doi:10.1016/j.ejmech.2009.11.019 PMID:19959260

Garro Martinez, J. C., Duchowicz, P. R., Estrada, M. R., Zamarbide, G. N., & Castro, E. A. (2011). QSAR study and molecular design of open-chain enaminones as anticonvulsant agents. *International Journal of Molecular Sciences*, *12*(12), 9354–9368. doi:10.3390/ijms12129354 PMID:22272137

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

Gasteiger, J., & Zupan, J. (1993). Neural networks in chemistry. *Angewandte Chemie*, *32*(4), 503–527. doi:10.1002/anie.199305031

Gaudin, C., Cunha, D., Ivanoff, E., Horcajada, P., Chevé, G., & Yasri, A. et al. (2012). A quantitative structure activity relationship approach to probe the influence of the functionalization on the drug encapsulation of porous metal-organic frameworks. *Microporous and Mesoporous Materials*, *157*, 124–130. doi:10.1016/j.micromeso.2011.06.011

Gázquez, J. L., Cedillo, A., & Vela, A. (2007). Electrodonating and electroaccepting powers. *The Journal of Physical Chemistry A*, *111*(10), 1966–1970. doi:10.1021/jp065459f PMID:17305319

Geçen, N., Sarıpınar, E., Yanmaz, E., & Şahin, K. (2012). Application of electron conformational-genetic algorithm approach to 1,4-dihydropyridines as calcium channel antagonists: Pharmacophore identification and bioactivity prediction. *Journal of Molecular Modeling*, *18*(1), 65–82. doi:10.1007/s00894-011-1024-5 PMID:21451963

Gedeck, P., & Lewis, R. A. (2008). Exploiting QSAR models in lead optimization. *Current Opinion in Drug Discovery & Development*, *11*(4), 569–575. PMID:18600573

Geerlings, P., De. Proft, F., & Langenaeker, W. (2003). Conceptual density functional theory. *Chemical Reviews*, *103*(5), 1793–1874. doi:10.1021/cr990029p PMID:12744694

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

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

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

George, S., Xia, T., Rallo, R., Zhao, Y., Ji, Z., & Lin, S. et al. (2011). Use of a high-throughput screening approach coupled with in vivo zebrafish embryo screening to develop hazard ranking for engineered nanomaterials. *ACS Nano*, *5*(3), 1805–1817. doi:10.1021/nn102734s PMID:21323332

German, N., Kaatz, G. W., & Kerns, R. J. (2008). Synthesis and evaluation of PSSRI-based inhibitors of *Staphylococcus aureus* multidrug efflux pumps. *Bioorganic & Medicinal Chemistry Letters*, *18*(4), 1368–1373. doi:10.1016/j.bmcl.2008.01.014 PMID:18242086

Ghanty, T. K., & Ghosh, S. K. (1993). Correlation between hardness, polarizability, and size of atoms, molecules, and clusters. *Journal of Physical Chemistry*, *97*(19), 4951–4953. doi:10.1021/j100121a015

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Ghanty, T. K., & Ghosh, S. K. (1996). A density functional approach to hardness, polarizability, and valency of molecules in chemical reactions. *Journal of Physical Chemistry*, *100*(30), 12295–12298. doi:10.1021/jp960276m

Ghasemi, J., & Saaidpour, S. (2007). QSPR prediction of aqueous solubility of drug-like organic compounds. *Chemical & Pharmaceutical Bulletin*, *55*(4), 669–674. doi:10.1248/cpb.55.669 PMID:17409570

Ghorbanzadeh, M., Fatemi, M. H., & Karimpour, M. (2012). Modeling the cellular uptake of magnetofluorescent nanoparticles in pancreatic cancer cells: A quantitative structure activity relationship study. *Industrial & Engineering Chemistry Research*, *51*(32), 10712–10718. doi:10.1021/ie3006947

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

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

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

Ghosh, S. K., & Berkowitz, M. J. (1985). A classical fluid-like approach to the density-functional formalism of many-electron systems. *The Journal of Chemical Physics*, *83*(6), 2976–2983. doi:10.1063/1.449846

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

Gironés, X., & Carbó-Dorca, R. (2006). Modelling Toxicity using Molecular Quantum Similarity Measures. *QSAR & Combinatorial Science*, *25*(7), 579–589. doi:10.1002/qsar.200530128

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

Gobas, F. A., de Wolf, W., Burkhard, L. P., Verbruggen, E., & Plotzke, K. (2009). Revisiting bioaccumulation criteria for POPs and PBT assessments. *Integrated Environmental Assessment and Management*, *5*(4), 624–637. doi:10.1897/IEAM_2008-089.1 PMID:19552497

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

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

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

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

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

Gombar, V. K., Mattioni, B. E., Zwickl, C., & Deahl, J. T. (2006) Computational approaches for assessment of toxicity: A historical perspective and current status. In Computational toxicology: Risk assessment for pharmaceutical and environmental chemicals. John Wiley & Sons, Inc. doi: (Chapter 7). doi:10.1002/9780470145890

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

Gong, N., Shao, K., Feng, W., Lin, Z., Liang, C., & Sun, Y. (2011). Biotoxicity of nickel oxide nanoparticles and bio-remediation by microalgae Chlorella vulgaris. *Chemosphere*, *83*(4), 510–516. doi:10.1016/j.chemosphere.2010.12.059 PMID:21216429

Gonzalez-Diaz, H., Arrasate, S., Gomez-Sanjuan, A., Sotomayor, N., Lete, E., Besada-Porto, L., & Ruso, J. M. (2013). General theory for multiple input-output perturbations in complex molecular systems. 1. Linear QSPR electronegativity models in physical, organic, and medicinal chemistry. *Current Topics in Medicinal Chemistry*, *13*(14), 1713–1741. doi:10.2174/1568026611313140011 PMID:23889050

Gonzalez-Diaz, H., Herrera-Ibata, D. M., Duardo-Sanchez, A., Munteanu, C. R., Orbegozo-Medina, R. A., & Pazos, A. (2014). ANN multiscale model of anti-HIV drugs activity vs AIDS prevalence in the US at county level based on information indices of molecular graphs and social networks. *Journal of Chemical Information and Modeling*, *54*(3), 744–755. doi:10.1021/ci400716y PMID:24521170

Gonzalez-Díaz, H., Perez-Bello, A., Cruz-Monteagudo, M., Gonzalez-Diaz, Y., Santana, L., & Uriarte, E. (2007). Chemometrics for QSAR with low sequence homology: Mycobacterial promoter sequences recognition with 2D-RNA entropies. *Chemometrics and Intelligent Laboratory Systems*, *85*(1), 20–26. doi:10.1016/j.chemolab.2006.03.005

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

González, M. P., Terán, C., Teijeira, M., & González-Moa, M. J. (2005). GETAWAY descriptors to predictingA2A adenosine receptors agonists. *European Journal of Medicinal Chemistry*, *40*(11), 1080–1086. doi:10.1016/j.ejmech.2005.04.014 PMID:16006015

Goodarzi, M., da Cunha, E. F. F., Freitas, M. P., & Ramalho, T. C. (2010). QSAR and docking studies of novel antileishmanial diaryl sulfides and sulfonamides. *European Journal of Medicinal Chemistry*, *45*(11), 4879–4889. doi:10.1016/j.ejmech.2010.07.060 PMID:20728249

Goodarzi, M., & Freitas, M. P. (2009). Prediction of electrophoretic enantioseparation of aromatic amino acids/esters through MIA-QSPR. *Separation and Purification Technology*, *68*(3), 363–366. doi:10.1016/j.seppur.2009.06.005

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., Freitas, M. P., & Ferreira, E. B. (2009). Influence of changes in 2-D chemical structure drawings and image formats on the prediction of biological properties using MIA-QSAR. *QSAR & Combinatorial Science*, *28*(4), 458–464. doi:10.1002/qsar.200810146

Goodarzi, M., Freitas, M. P., & Jensen, R. (2009). Ant colony optimization as a feature selection method in the QSAR modeling of anti-HIV-1 activities of 3-(3,5-dimethylbenzyl)uracil derivatives using MLR, PLS and SVM regressions. *Chemometrics and Intelligent Laboratory Systems*, *98*(2), 123–129. doi:10.1016/j.chemolab.2009.05.005

Goodarzi, M., Freitas, M. P., & Jensen, R. (2009). Feature selection and linear/nonlinear regression methods for the accurate prediction of glycogen synthase kinase-3β inhibitory activities. *Journal of Chemical Information and Modeling*, *49*(4), 824–832. doi:10.1021/ci9000103 PMID:19338295

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

Gordon, M., & Scantlebury, G. R. (1964). Non-random polycondensation: Statistical theory of the substitution effect. *Transactions of the Faraday Society*, *60*, 604–621. doi:10.1039/tf9646000604

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Gordon, T., Chen, L. C., Fine, J. M., Schlesinger, R. B., Su, W. Y., Kimmel, T. A., & Amdur, M. O. (1992). Pulmonary effects of inhaled zinc oxide in human subjects, guinea pigs, rats, and rabbits. *American Industrial Hygiene Association Journal*, *53*(8), 503–509. doi:10.1080/15298669291360030 PMID:1509990

Gottschalk, F., Sun, T. Y., & Nowack, B. (2013). Environmental concentrations of engineered nanomaterials: Review of modelling and analytical studies. *Environmental Pollution*, *181*, 287–300. doi:10.1016/j.envpol.2013.06.003 PMID:23856352

Gouin, T., Mackay, D., Jones, K. C., Harner, T., & Meijer, S. N. (2004). Evidence for the "grasshopper" effect and fractionation during long-range atmospheric transport of organic contaminants. *Environmental Pollution*, *128*(1-2), 139–148. doi:10.1016/j.envpol.2003.08.025 PMID:14667724

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

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

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

Gramatica, P. (2010). Chemometric methods and theoretical molecular descriptors in predictive QSAR modeling of the environmental behavior of organic pollutants. In T. Puzyn, J. Leszczynski, & M. T. D. Cronin (Eds.), *Recent advances in QSAR studies. Challenges and advances in computational chemistry and physics*. Dordrecht, The Netherlands: Springer. doi:10.1007/978-1-4020-9783-6_12

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

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

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

Griffith, D. J. (1999). *Introduction to electrodynamics*. Prentice Hall, Inc.

Griffitt, R. J., Luo, J., Gao, J., Bonzongo, J. C., & Barber, D. S. (2008). Effects of particle composition and species on toxicity of metallic nanomaterials in aquatic organisms. *Environmental Toxicology and Chemistry*, *27*(9), 1972–1978. doi:10.1897/08-002.1 PMID:18690762

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

Grover, M., Singh, B., Bakshi, M., & Singh, S. (2000a). Quantitative structure–Property relationships in pharmaceutical research – Part 1. *Pharmaceutical Science & Technology Today*, *3*, 28–35.

Grover, M., Singh, B., Bakshi, M., & Singh, S. (2000b). Quantitative structure–property relationships in pharmaceutical research – Part 2. *Pharmaceutical Science & Technology Today*, *3*(2), 50–57. doi:10.1016/S1461-5347(99)00215-1 PMID:10664573

EBSCOhost - printed on 2/14/2023 7:16 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

Guha, R., & Jurs, P. C. (2005). Determining the validity of a QSAR model-a classification approach. *Journal of Chemical Information and Modeling*, *45*(1), 65–73. doi:10.1021/ci0497511 PMID:15667130

Guimarães, M. C., Silva, D. G., da Mota, E. G., da Cunha, E. F. F., & Freitas, M. P. (2014). Computer-assisted design of dual-target anti-HIV-1 compounds. *Medicinal Chemistry Research*, *23*(3), 1548–1558. doi:10.1007/s00044-013-0765-3

Guner, O. F. (2000). *Pharmacophore perception, development, and use in drug design*. La Jolla, CA: International University Line.

Gunn, S. R. (1997). *Support vector machines for classification and regression*. University of Southampton.

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.

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

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

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.

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

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

Gutman, I. (1988). Topological analysis of Eigenvalues of the adjacency matrices in graph theory: A difficulty with the concept of internal connectivity. *Chemical Physics Letters*, *148*(1), 93–94. doi:10.1016/0009-2614(88)87267-1

Gutman, I. (1997). Permanents of adjacency matrices and their dependence on molecular structure. *Polycyclic Aromatic Compounds*, *12*(4), 281–287. doi:10.1080/10406639808233845

Gutman, I., & Trinajstić, N. (1972). Graph theory and molecular orbitals. Total π-electron energy of alternant hydrocarbons. *Chemical Physics Letters*, *17*(4), 535–538. doi:10.1016/0009-2614(72)85099-1

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

Gyftopoulos, E. P., & Hatsopoulos, G. N. (1968). Quantum-thermodynamic definition of electronegativity. *Proceedings of the National Academy of Sciences of the United States of America*, *60*(3), 786–793. doi:10.1073/pnas.60.3.786 PMID:16591659

Haase, A., Tentschert, J., & Luch, A. (2012). Nanomaterials: A challenge for toxicological risk assessment? In A. Luch (Ed.), *Molecular, clinical and environmental toxicology: Environmental toxicology* (pp. 219–250). Heidelberg: Springer. doi:10.1007/978-3-7643-8340-4_8

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Haghdadi, M., & Fatemi, M. H. (2010). Quantitative structure–Toxicity prediction of log(1/EC50) for some benzene derivatives from their density functional theory calculated molecular descriptors. *Bulletin of the Chemical Society of Japan*, *83*(4), 345–350. doi:10.1246/bcsj.20090309

Hair, J. F. Jr, Anderson, R. E., Tatham, R. L., & Black, W. C. (2005). *Multivariate data analysis*. Singapore: Pearson Education.

Halling-Sorensen, B., Nors Nielsen, S., Lanzky, P. F., Ingerslev, F., Holten Lutzhoft, H. C., & Jorgensen, S. E. (1998). Occurrence, fate and effects of pharmaceutical substances in the environment--a review. *Chemosphere*, *36*(2), 357–393. doi:10.1016/S0045-6535(97)00354-8 PMID:9569937

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. (1995). Electrotopological state indices for atom types: A novel combination of electronic, topological, and valence state information. *Journal of Chemical Information and Computer Sciences*, *35*(6), 1039–1045. doi:10.1021/ci00028a014

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

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

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

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

Hanczar, B., Hua, J., Sima, C., Weinstein, J., Bittner, M., & Dougherty, E. R. (2010). Small-sample precision of ROC-related estimates. *Bioinformatics (Oxford, England)*, *26*(6), 822–830. doi:10.1093/bioinformatics/btq037 PMID:20130029

Handy, R. D., & Shaw, B. J. (2007). Toxic effects of nanoparticles and nanomaterials: Implications for public health, risk assessment and the public perception of nanotechnology. *Health Risk & Society*, *9*(2), 125–144. doi:10.1080/13698570701306807

Handzlik, J., Matys, A., & Kieć-Kononowicz, K. (2013). Recent advances in multi-drug resistance (MDR) efflux pump inhibitors of Gram-positive bacteria *S. aureus*. *Antibiotics*, *2*(1), 28–45. doi:10.3390/antibiotics2010028

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

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

Hannongbua, S., Nivasanond, K., Lawtrakul, L., Pungpo, P., & Wolschann, P. (2001). 3D-Quantitative structure−activity relationships of HEPT derivatives as HIV-1 reverse transcriptase inhibitors, based on ab initio calculations. *Journal of Chemical Information and Computer Sciences*, *41*(3), 848–855. doi:10.1021/ci0001278 PMID:11410067

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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. (1993). Quantitative structure-activity relationships and the unnamed science. *Accounts of Chemical Research*, *26*(4), 147–153. doi:10.1021/ar00028a003

Hansch, C., & Clayton, J. M. (1973). Lipophilic character and biological activity of drugs II: The parabolic case. *Journal of Pharmaceutical Sciences*, *62*(1), 1–21. doi:10.1002/jps.2600620102 PMID:4566530

Hansch, C., & 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

Hansch, C., & Fujita, T. (1995). *Classical and three-dimensional QSAR in agrochemistry*. ACS Publications. doi:10.1021/bk-1995-0606

Hansch, C., & Gao, H. (1997). Comparative QSAR: Radical reactions of benzene derivatives in chemistry and biology. *Chemical Reviews*, *97*(8), 2995–3060. doi:10.1021/cr9601021 PMID:11851484

Hansch, C., Gao, H., & Hoekman, D. (1998). *A generalized approach to comparative QSAR*. Washington, DC: Taylor and Francis.

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., Kim, D., Leo, A. J., Novellino, E., Silipo, C., & Vittoria, A. et al. (1989). Toward a quantitative comparative toxicology of organic compounds. *CRC Critical Reviews in Toxicology*, *19*(3), 185–226. doi:10.3109/10408448909037471 PMID:2653732

Hansch, C., & Klein, T. E. (1986). Molecular graphics and QSAR in the study of enzyme-ligand interactions. On the definition of bioreceptors. *Accounts of Chemical Research*, *19*(12), 392–400. doi:10.1021/ar00132a003

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. (1995). *Exploring QSAR: fundamentals and applications in chemistry and biology*. Washington, DC: American Chemical Society.

Hansch, C., Leo, A., Hoekman, D., & Heller, S. R. (1995). *Exploring Qsar*. Washington, DC: American Chemical Society.

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., & Steward, A. R. (1964). The use of substituent constants in the analysis of the structure–activity relationship in penicillin derivatives. *Journal of Medicinal Chemistry*, *7*(6), 691–694. doi:10.1021/jm00336a001 PMID:14262788

Hansen, C., Telzer, B. R., & Zhang, L. (1995). Comparative QSAR in toxicology: Examples from teratology and cancer chemotherapy of aniline mustards. *Critical Reviews in Toxicology*, *25*(1), 67–89. doi:10.3109/10408449509089887 PMID:7734060

Hansen, N. T., Kouskoumvekaki, I., Jørgensen, F. S., Brunak, S., & Jonsdottir, S. O. (2006). Prediction of pH-dependent aqueous solubility of druglike molecules. *Journal of Chemical Information and Modeling*, *46*(6), 2601–2609. doi:10.1021/ci600292q PMID:17125200

Hansen, S. F., Larsen, B. H., Olsen, S. I., & Baun, A. (2007). Categorization framework to aid hazard identification of nanomaterials. *Nanotoxicology*, *1*(3), 243–250. doi:10.1080/17435390701727509

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

Haranczyk, M., Puzyn, T., & Sadowski, P. (2008). CongGENER - A tool for modeling of the congeneric sets of environmental pollutants. *QSAR & Combinatorial Science*, *27*(7), 826–833. doi:10.1002/qsar.200710149

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

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

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

Hartmann, R. W., Bayer, H., Grun, G., Sergejew, T., Bartz, U., & Mitrenga, M. (1995). Pyridyl-substituted tetrahydr ocyclopropa[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

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

Hassellov, R., & Kaegi, R. (2009). Analysis and characterization of manufactured nanoparticles in aquatic environmants. In Nanoscience and nanotechnology: Environmental and human health implications. Wiley. doi:10.1002/9781444307504.ch6

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

Heinlaan, M., Ivask, A., Blinova, I., Dubourguier, H. C., & Kahru, A. (2008). Toxicity of nanosized and bulk ZnO, CuO and TiO_2 to bacteria Vibrio fischeri and crustaceans Daphnia magna and Thamnocephalus platyurus. *Chemosphere*, *71*(7), 1308–1316. doi:10.1016/j.chemosphere.2007.11.047 PMID:18194809

Helguera, A. M., Cabrera Perez, M. A., Gonzalez, M. P., Ruiz, R. M., & Gonzalez Diaz, H. (2005). A topological substructural approach applied to the computational prediction of rodent carcinogenicity. *Bioorganic & Medicinal Chemistry*, *13*(7), 2477–2488. doi:10.1016/j.bmc.2005.01.035 PMID:15755650

Helguera, A. M., Cordeiro, M. N. D. S., Pérez, M. A. C., Combes, R. D., & González, M. P. (2008). QSAR modeling of the rodent carcinogenicity of nitrocompounds. *Bioorganic & Medicinal Chemistry*, *16*(6), 3395–3407. doi:10.1016/j.bmc.2007.11.029 PMID:18295489

Helguera, A. M., Gonzalez, M. P., Cordeiro, M. N. D. S., & Perez, M. A. (2007). Quantitative structure carcinogenicity relationship for detecting structural alerts in nitroso-compounds. *Toxicology and Applied Pharmacology*, *221*(2), 189–202. doi:10.1016/j.taap.2007.02.021 PMID:17477948

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Hemalatha, T., Imran, P. K. M., Gnanamani, A., & Nagarajan, S. (2012). QSAR and evaluation of molecular electrostatic potential for N-nitrosopiperidinone semicarbazones. *Chemometrics and Intelligent Laboratory Systems*, *116*, 87–93. doi:10.1016/j.chemolab.2012.04.014

Hemmateenejad, B. (2004). Optimal QSAR analysis of the carcinogenic activity of drugs by correlation ranking and genetic algorithm-based PCR. *Journal of Chemometrics*, *18*(11), 475–485. doi:10.1002/cem.891

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

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

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

He, X., Szewczyk, P., Karyakin, A., Evin, M., Hong, W.-X., Zhang, Q., & Chang, G. (2010). Structure of a cation-bound multidrug and toxic compound extrusion transporter. *Nature*, *467*(7318), 991–994. doi:10.1038/nature09408 PMID:20861838

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

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

Higgins, C. F. (2001). ABC transporters: Physiology, structure and mechanism–an overview. *Research in Microbiology*, *152*(3), 205–210. doi:10.1016/S0923-2508(01)01193-7 PMID:11421269

Higgins, W. A. (1789). *Comparative view of the phlogistic and antiphlogistic theories*. London: J. Murray.

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

Hilal, R., & Elroby, S. A. K. (2011). A QSAR study for 2-(4-aminophenyl)benzothiazoles: Using DFT optimisation of geometry of molecules. *Molecular Simulation*, *37*(1), 62–71. doi:10.1080/08927022.2010.520133

Hilal, S. H., Carreira, L. A., & Karickhoff, S. W. (1994). Estimation of chemical reactivity parameter and physical properties of organic molecules using SPARC. In P. Politzer & J. S. Murray (Eds.), *Quantitative treatments of solute/solvent interactions* (pp. 291–353). Amsterdam: Elsevier.

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

Hill, T., & Lewicki, P. (2006). *STATISTICS methods and applications. A comprehensive reference for science, industry and data mining*. Tulsa, OK: StatSoft.

Hinchliffe, P., Greene, N. P., Paterson, N. G., Crow, A., Hughes, C., & Koronakis, V. (2014). Structure of the periplasmic adaptor protein from a major facilitator superfamily (MFS) multidrug efflux pump. *FEBS Letters*, *588*(17), 3147–3153. doi:10.1016/j.febslet.2014.06.055 PMID:24996185

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

Hohenberg, P., & Kohn, W. (1964). Inhomogeneous electron gas. *Physical Review B: Condensed Matter and Materials Physics*, *136*(3B), B864–B871. doi:10.1103/PhysRev.136.B864

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Holden, P. A., Nisbet, R. M., Lenihan, H. S., Miller, R. J., Cherr, G. N., Schimel, J. P., & Gardea-Torresdey, J. L. (2013). Ecological nanotoxicology: Integrating nanomaterial hazard considerations across the subcellular, population, community, and ecosystems levels. *Accounts of Chemical Research*, *46*(3), 813–822. doi:10.1021/ar300069t PMID:23039211

Hollander, A., Scheringer, M., Shatalov, V., Mantseva, E., Sweetman, A., Roemer, M., & van de Meent, D. et al. (2008). Estimating overall persistence and long-range transport potential of persistent organic pollutants. a comparison of seven multimedia mass balance models and atmospheric transport models. *Journal of Environmental Monitoring*, *10*(10), 1139–1147. doi:10.1039/b803760d PMID:18843390

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

Holland, J. H. (1992). *Adaptiation in natural and atrificials systems*. MIT Press.

Hollenstein, K., Dawson, R. J., & Locher, K. P. (2007). Structure and mechanism of ABC transporter proteins. *Current Opinion in Structural Biology*, *17*(4), 412–418. doi:10.1016/j.sbi.2007.07.003 PMID:17723295

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

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

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

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

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 Medicinal Chemistry*, *119*, 10509–10524.

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

Hosoya, H. (1971). Topological index: A newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons. *Bulletin of the Chemical Society of Japan*, *44*(9), 2332–2339. doi:10.1246/bcsj.44.2332

Hosoya, H. (1972). Topological index as a sorting device for coding chemical structures. *Journal of Chemical Documentation*, *12*(3), 181–183. doi:10.1021/c160046a010

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

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

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

Hsu, D. D. (2013). *Chemicool periodic table*. Retrieved from http://www.chemicool.com/

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

Hughes, M. F. (2002). Arsenic toxicity and potential mechanisms of action. *Toxicology Letters*, *133*(1), 1–16. doi:10.1016/S0378-4274(02)00084-X PMID:12076506

Hund-Rinke, K., & Simon, M. (2006). Ecotoxic effect of photocatalytic active nanoparticles (TiO2) on algae and daphnids. *Environmental Science and Pollution Research International*, *13*(4), 225–232. doi:10.1065/espr2006.06.311 PMID:16910119

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.

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

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

Hutson, N. D., Ryan, S. P., & Touati, A. (2009). Assessment of PCDD/F and PBDD/F emissions from coal-fired power plants during injection of brominated activated carbon for mercury control. *Atmospheric Environment*, *43*(26), 3973–3980. doi:10.1016/j.atmosenv.2009.05.026

Huuskonen, J. (2000). Estimation of aqueous solubility for a diverse set of organic compounds based on molecular topology. *Journal of Chemical Information and Computer Sciences*, *40*(3), 773–777. doi:10.1021/ci9901338 PMID:10850781

Hvorup, R. N., Winnen, B., Chang, A. B., Jiang, Y., Zhou, X. F., & Saier, M. H. Jr. (2003). The multidrug/oligosaccharidyl-lipid/polysaccharide (MOP) exporter superfamily. *European Journal of Biochemistry*, *270*(5), 799–813. doi:10.1046/j.1432-1033.2003.03418.x PMID:12603313

Ibezim, E., Duchowicz, P. R., Ortiz, E. V., & Castro, E. A. (2012). QSAR on aryl-piperazine derivatives with activity on malaria. *Chemometrics and Intelligent Laboratory Systems*, *110*(1), 81–88. doi:10.1016/j.chemolab.2011.10.002

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

Inoko, H., & Yoshihara, E. (2011). *7985410*. Washington, DC: US Patent Office.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Ispas, C., Andreescu, D., Patel, A., Goia, D. V., Andreescu, S., & Wallace, K. N. (2009). Toxicity and developmental defects of different sizes and shape nickel nanoparticles in zebrafish. *Environmental Science & Technology*, *43*(16), 6349–6356. doi:10.1021/es9010543 PMID:19746736

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

Ivanciuc, O., Ivanciuc, T., & Balaban, A. T. (1998). Quantitative structure-property relationship study of normal boiling points for halogen-/oxygen-/sulfur-containing organic compounds using the CODESSA program. *Tetrahedron*, *54*(31), 9129–9142. doi:10.1016/S0040-4020(98)00550-X

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

Jack, D. L., Yang, N. M., & Saier, H, M. (. (2001). The drug/metabolite transporter superfamily. *European Journal of Biochemistry*, *268*(13), 3620–3639. doi:10.1046/j.1432-1327.2001.02265.x PMID:11432728

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

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

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. (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. (2006). Scoring functions for protein-ligand docking. *Current Protein & Peptide Science*, *7*(5), 407–420. doi:10.2174/138920306778559395 PMID:17073693

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

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

Jana, G. A., Delgado, E. J., & Medina, F. E. (2014). How important is the synclinal conformation of sulfonylureas to explain the inhibition of AHAS: A theoretical study. *Journal of Chemical Information and Modeling*, *54*(3), 926–932. doi:10.1021/ci400721y PMID:24548139

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Jarmuła, A., Obłąk, E., Wawrzycka, D., & Gutowicz, J. (2010). Oporność wielolekowa związana z aktywnym usuwaniem leków z komórek drobnoustrojów. *Postepy higieny i medycyny doswiadczalnej (Online)*, *65*, 216–227. doi:10.5604/17322693.937011 PMID:21502698

Jaworska, J. S., Comber, M., Auer, C., & Van Leeuwen, C. J. (2003). Summary of a workshop on regulatory acceptance of (Q)SARs for human health and environmental endpoints. *Environmental Health Perspectives*, *111*(10), 1358–1360. doi:10.1289/ehp.5757 PMID:12896859

Jaworska, J. S., Nikolova-Jeliazkova, N., & Aldenberg, T. (2004). *Review of methods for applicability domain estimation. Report*. Ispra, Italy: The European Commission-Joint Research Centre.

Jaworska, J., Nikolova-Jeliazkova, N., & Aldenberg, T. (2005). QSAR applicability domain estimation by projection of the training set descriptor space: A review. *Alternatives to Laboratory Animals*, *33*(5), 445–459. PMID:16268757

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

Jayatilleke, P. R. N., Nair, A. C., Zauhar, R., & Welsh, W. J. (2000). Computational studies on HIV-1 protease inhibitors: Influence of calculated inhibitor−enzyme binding affinities on the statistical quality of 3D-QSAR CoMFA models. *Journal of Medicinal Chemistry*, *43*(23), 4446–4451. doi:10.1021/jm9905357 PMID:11087569

Jensen, R. (2006). *Studies in computational intelligence* (A. Abraham, C. Grosan, & V. Ramos, Eds.). Heidelberg: Springer.

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

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

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

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

Ji, Z., Wang, X., Zhang, H., Lin, S., Meng, H., & Sun, B. et al. (2012). Designed synthesis of CeO_2 nanorods and nanowires for studying toxicological effects of high aspect ratio nanomaterials. *ACS Nano*, *6*(6), 5366–5380. doi:10.1021/nn3012114 PMID:22564147

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

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

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

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.

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

Jouan-Rimbaud, D., Bouveresse, E., Massart, D. L., & de Noord, O. E. (1999). Detection of prediction outliers and inliers in multivariate calibration. *Analytica Chimica Acta*, *388*(3), 283–301. doi:10.1016/S0003-2670(98)00626-6

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

Kaatz, G. W., Moudgal, V. V., Seo, S. M., Hansen, J. B., & Kristiansen, J. E. (2003). Phenylpiperidine selective serotonin reuptake inhibitors interfere with multidrug efflux pump activity in *Staphylococcus aureus. International Journal of Antimicrobial Agents*, *22*(3), 254–261. doi:10.1016/S0924-8579(03)00220-6 PMID:13678830

Kaatz, G. W., Moudgal, V. V., Seo, S. M., & Kristiansen, J. E. (2003). Phenothiazines and thioxanthenes inhibit multidrug efflux pump activity in *Staphylococcus aureus. Antimicrobial Agents and Chemotherapy*, *47*(2), 719–726. doi:10.1128/AAC.47.2.719-726.2003 PMID:12543683

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

Kahru, A., & Ivask, A. (2013). Mapping the dawn of nanoecotoxicological research. *Accounts of Chemical Research*, *46*(3), 823–833. doi:10.1021/ar3000212 PMID:23148404

Kalia, N. P., Mahajan, P., Mehra, R., Nargotra, A., Sharma, J. P., Koul, S., & Khan, I. A. (2012). Capsaicin, a novel inhibitor of the NorA efflux pump, reduces the intracellular invasion of Staphylococcus aureus. *The Journal of Antimicrobial Chemotherapy*, *67*(10), 2401–2408. doi:10.1093/jac/dks232 PMID:22807321

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

Kamlet, M. J., Abboud, J.-L. M., Abraham, M. H., & Taft, R. W. (1983). Linear solvation energy relationships. 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

Kang, S. J., Kim, B. M., Lee, Y. J., & Chung, H. W. (2008). Titanium dioxide nanoparticles trigger P53-mediated damage response in peripheral blood lymphocytes. *Environmental and Molecular Mutagenesis*, *49*(5), 399–405. doi:10.1002/em.20399 PMID:18418868

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

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

Karabunarliev, S., Mekenyan, O. G., Karcher, W., Russom, C. L., & Bradbury, S. P. (1996). Quantum-chemical descriptors for estimating the acute Ttoxicity of substituted benzenes to the Guppy (*Poecilia reticulata*) and Fathead Minnow (*Pimephales promelas*). *Quantitative Structure-Activity Relationships*, *15*(4), 311–320. doi:10.1002/qsar.19960150406

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Karcher, W., & Devillers, J. (1990). Practical applications of quantitative structure-activity relationships (QSAR). *Environmental Chemistry and Toxicology, 1*.

Karelson, M. (2000). *Molecular descriptors in QSAR/QSPR*. New York: Wiley-InterScience.

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

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

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

Karlsson, H. L., Cronholm, P., Gustafsson, J., & Möller, L. (2008). Copper oxide nanoparticles are highly toxic: A comparison between metal oxide nanoparticles and carbon nanotubes. *Chemical Research in Toxicology*, *21*(9), 1726–1732. doi:10.1021/tx800064j PMID:18710264

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

Kar, S. K., & Roy, K. (2011). Development and validation of a robust QSAR model for prediction of carcinogenicity of drugs. *Indian Journal of Biochemistry & Biophysics*, *48*, 111–122. PMID:21682143

Kar, S., Gajewicz, A., Puzyn, T., & Roy, K. (2014a). Nano-quantitative structure-activity relationship modeling using easily computable and interpretable descriptors for uptake of magnetofluorescent engineered nanoparticles in pancreatic cancer cells. *Toxicology in Vitro*, *28*(4), 600–606. doi:10.1016/j.tiv.2013.12.018 PMID:24412539

Kar, S., Gajewicz, A., Puzyn, T., Roy, K., & Leszczynski, J. (2014b). Periodic table-based descriptors to encode cytotoxicity profile of metal oxide nanoparticles: A mechanistic QSTR approach. *Ecotoxicology and Environmental Safety*, *107*, 162–169. doi:10.1016/j.ecoenv.2014.05.026 PMID:24949897

Kar, S., & Roy, K. (2010). Predictive toxicology using QSAR: A perspective. *Journal of the Indian Chemical Society*, *87*, 1455–1515.

Kar, S., & Roy, K. (2010). QSAR modeling of toxicity of diverse organic chemicals to Daphnia magna using 2D and 3D descriptors. *Journal of Hazardous Materials*, *177*(1-3), 344–351. doi:10.1016/j.jhazmat.2009.12.038 PMID:20045248

Kar, S., & Roy, K. (2012). First report on development of quantitative interspecies structure–carcinogenicity relationship models and exploring discriminatory features for rodent carcinogenicity of diverse organic chemicals using OECD guidelines. *Chemosphere*, *87*(4), 339–355. doi:10.1016/j.chemosphere.2011.12.019 PMID:22225702

Kasemets, K., Ivask, A., Dubourguier, H. C., & Kahru, A. (2009). Toxicity of nanoparticles of ZnO, CuO and TiO_2 to yeast Saccharomyces cerevisiae. *Toxicology In Vitro*, *23*(6), 1116–1122. doi:10.1016/j.tiv.2009.05.015 PMID:19486936

Katritzky, A. R., Lobanov, V. S., & Karelson, M. (1994). CODESSA, reference manual. University of Florida. Retrieved from http://www.semichem.com/codessa

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Katritzky, A. R., Fara, D. C., Petrukhin, R. O., Tatham, D. B., Maran, U., Lomaka, A., & Karelson, M. (2002). The present utility and future potential for medicinal chemistry of QSAR / QSPR with whole molecule descriptors. *Current Topics in Medicinal Chemistry*, *2*(12), 1333–1356. doi:10.2174/1568026023392922 PMID:12470284

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

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

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.

Katritzky, A. R., Maran, U., Lobanov, V. S., & Karelson, M. (2000). Structurally diverse quantitative structure-property relationship correlations of technologically relevant physical properties. *Journal of Chemical Information and Computer Sciences*, *40*(1), 1–18. doi:10.1021/ci9903206 PMID:10661545

Katritzky, A. R., Mu, L., Lobanov, V. S., & Karelson, M. (1996). Correlation of boiling points with molecular structure. A training set of 298 diverse organics and a test set of 9 simple inorganics. *Journal of Physical Chemistry*, *100*(24), 10400–10407. doi:10.1021/jp953224q

Katritzky, A. R., Petrukhin, R., Yang, H., & Karelson, M. (2005). *CODESSA PRO*. Florida, USA.

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

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

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

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

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

Kern, W. V., Steinke, P., Schumacher, A., Schuster, S., von Baum, H., & Bohnert, J. A. (2006). Effect of 1-(1-naphthylmethyl)-piperazine, a novel putative efflux pump inhibitor, on antimicrobial drug susceptibility in clinical isolates of Escherichia coli. *The Journal of Antimicrobial Chemotherapy*, *57*(2), 339–343. doi:10.1093/jac/dki445 PMID:16354747

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

Khan, I. A., Mirza, Z. M., Kumar, A., Verma, V., & Qazi, G. N. (2006). Piperine, a phytochemical potentiator of ciprofloxacin against Staphylococcus aureus. *Antimicrobial Agents and Chemotherapy*, *50*(2), 810–812. doi:10.1128/AAC.50.2.810-812.2006 PMID:16436753

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

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

Kikukawa, T., Nara, T., Araiso, T., Miyauchi, S., & Kamo, N. (2006). Two-component bacterial multidrug transporter, EbrAB: Mutations making each component solely functional. *Biomembranes*, *1758*(5), 673–679. doi:10.1016/j.bbamem.2006.04.004

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

Kim, K. H. (1995). Molecular similarity. In P. M. Dean (Ed.), *Drug design* (pp. 291–331). New York: Chapman & Hall; doi:10.1007/978-94-011-1350-2_12

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

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

Kim, Y., Dalhaimer, P., Christian, D. A., & Discher, D. E. (2005). Polymeric worm micelles as nano-carriers for drug delivery. *Nanotechnology*, *16*(7), S484–S491. doi:10.1088/0957-4484/16/7/024 PMID:21727469

Kiralj, R., & Ferreira, M. (2008). Comparative chemometric and QSAR/SAR study of structurally unrelated substrates of a MATE efflux pump VmrA from V. parahaemolyticus: Prediction of multidrug resistance. *QSAR & Combinatorial Science*, *27*(3), 314–329. doi:10.1002/qsar.200630164

Kirchhoff, G. R. (1847). Über die auflösung der gleichungen, auf welche man bei der untersuchung der linearen verteilung galvanischer ströme geführt wird. *Annales der Physik et Chimie*, *72*(12), 497–508. doi:10.1002/andp.18471481202

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.

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

Klaine, S. J., Alvarez, P. J. J., Batley, G. E., Fernandes, T. F., Handy, R. D., & Lyon, D. Y. et al. (2008). Nanomaterials in the environment: Behavior, fate, bioavailability, and effects. *Environmental Toxicology and Chemistry*, *27*(9), 1825–1851. doi:10.1897/08-090.1 PMID:19086204

Klasmeier, J., Matthies, M., Macleod, M., Fenner, K., Scheringer, M., Stroebe, M., & Wania, F. et al. (2006). Application of multimedia models for screening assessment of long-range transport potential and overall persistence. *Environmental Science & Technology*, *40*(1), 53–60. doi:10.1021/es0512024 PMID:16433332

Kleandrova, V. V., Luan, F., Gonzalez-Diaz, H., Ruso, J. M., Melo, A., Speck-Planche, A., & Cordeiro, M. N. (2014). Computational ecotoxicology: Simultaneous prediction of ecotoxic effects of nanoparticles under different experimental conditions. *Environment International*, *73*, 288–294. doi:10.1016/j.envint.2014.08.009 PMID:25173945

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

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

Kochakian, C. D. (1976). *Anabolic-androgenic steroids*. Berlin: Springer. doi:10.1007/978-3-642-66353-6

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

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, *43*(1), 59–69. doi:10.1007/BF00337288

Kolaczkowski, M., Michalak, K., & Motohashi, N. (2003). Phenothiazines as potent modulators of yeast multidrug resistance. *International Journal of Antimicrobial Agents*, *22*(3), 279–283. doi:10.1016/S0924-8579(03)00214-0 PMID:13678835

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

Kormann, F., Rominger, J., Schowanek, D., Wagner, J.-O., Schroder, R., & Wind, T. et al.. (2006). Modelling the fate of down-the-drain chemicals in rivers: An improved software for GREAT-ER. *Environmental Modelling & Software*, *21*(7), 925–936. doi:10.1016/j.envsoft.2005.04.009

Koronakis, V., Sharff, A., Koronakis, E., Luisi, B., & Hughes, C. (2000). Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. *Nature*, *405*(6789), 914–919. doi:10.1038/35016007 PMID:10879525

Koul, S., Koul, J. L., Taneja, S. C., Dhar, K. L., Jamwal, D. S., & Singh, K. et al. (2000). Structure–activity relationship of piperine and its synthetic analogues for their inhibitory potentials of rat hepatic microsomal constitutive and inducible cytochrome P450 activities. *Bioorganic & Medicinal Chemistry*, *8*(1), 251–268. doi:10.1016/S0968-0896(99)00273-4 PMID:10968285

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

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

Kruhlak, N. L., Contrera, J. F., Daniel Benz, R., & Matthews, E. J. (2007). Progress in QSAR toxicity screening of pharmaceutical impurities and other FDA regulated products. *Advanced Drug Delivery Reviews*, *59*(1), 43–55. doi:10.1016/j.addr.2006.10.008 PMID:17229485

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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. (1976). Quantitative structure-activity relationships, 4. Nonlinear dependence of biological activity on hydrophobic character: A new model. *Arzneimittel-Forschung. Drug Research*, *26*(11), 1991–1997. PMID:1037231

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

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

Kubinyi, H. (1998). Comparative molecular field analysis (CoMFA). In N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer III, & P. R. Schreiner (Eds.), *The encyclopedia of computational chemistry* (pp. 448–460). Chichester, UK: Wiley.

Kubinyi, H. (Ed.). (1993). *3D QSAR in drug design: Theory, methods and applications*. Leiden, The Netherlands: ESCOM.

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

Kulkarni, S. A., & Zhu, J. (2008). Integrated approach to assess the domain of applicability of some commercial (Q)SAR models. *SAR and QSAR in Environmental Research*, *19*(1-2), 39–54. doi:10.1080/10629360701843508 PMID:18311633

Kumar, A., Khan, I. A., Koul, S., Koul, J. L., Taneja, S. C., & Ali, I. et al. (2008). Novel structural analogues of piperine as inhibitors of the NorA efflux pump of *Staphylococcus aureus*. *The Journal of Antimicrobial Chemotherapy*, *61*(6), 1270–1276. doi:10.1093/jac/dkn088 PMID:18334493

Kumar, D., Singh, R. N., Sahu, S., & Baboo, V. (2011). Hardness based quantitative structure toxicity relationship (QSTR) study on a series of aliphatic alcohol derivatives. *Organic Chemistry: An Indian Journal*, *7*, 41–47.

Kumar, S., & Varela, M. F. (2012). Biochemistry of bacterial multidrug efflux pumps. *International Journal of Molecular Sciences*, *13*(4), 4484–4495. doi:10.3390/ijms13044484 PMID:22605991

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

Kušić, H., Rasulev, B., Leszczynska, D., Leszczynski, J., & Koprivanac, N. (2009). Prediction of rate constants for radical degradation of aromatic pollutants in water matrix: A QSAR study. *Chemosphere*, *75*(8), 1128–1134. doi:10.1016/j.chemosphere.2009.01.019 PMID:19201442

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

La Du, B. N. (1994). Human cholinesterases and anticholinesterases. *American Journal of Human Genetics*, *55*, 593–594.

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

Langridge, R., & Klein, T. (1990). Molecular graphics and drug design. Comprehensive Medicinal Chemistry, 4, 413-429.

Lanone, S., Rogerieux, F., Geys, J., Dupont, A., Maillot-Marechal, E., & Boczkowski, J. et al. (2009). Comparative toxicity of 24 manufactured nanoparticles in human alveolar epithelial and macrophage cell lines. *Particle and Fibre Toxicology*, *6*(1), 14. doi:10.1186/1743-8977-6-14 PMID:19405955

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

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

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

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

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

Lee, A., Mao, W., Warren, M. S., Mistry, A., Hoshino, K., & Okumura, R. et al. (2000). Interplay between efflux pumps may provide either additive or multiplicative effects on drug resistance. *Journal of Bacteriology*, *182*(11), 3142–3150. doi:10.1128/JB.182.11.3142-3150.2000 PMID:10809693

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

Leermakers, M., Baeyens, W., De Gieter, M., Smedts, B., Meert, C., & De Bisschop, H. C. et al. (2006). Toxic arsenic compounds in environmental samples: Speciation and validation. *Trends in Analytical Chemistry*, *25*(1), 1–10. doi:10.1016/j.trac.2005.06.004

Lehtinen, J., & Lilius, E.-M. (2007). Promethazine renders *Escherichia coli* susceptible to penicillin G: Real-time measurement of bacterial susceptibility by fluoro-luminometry. *International Journal of Antimicrobial Agents*, *30*(1), 44–51. doi:10.1016/j.ijantimicag.2007.02.019 PMID:17475447

Lei, H. T., Shen, Z., Surana, P., Routh, M. D., Su, C. C., Zhang, Q., & Yu, E. W. (2011). Crystal structures of CmeR-bile acid complexes from Campylobacter jejuni. *Protein Science*, *20*(4), 712–723. doi:10.1002/pro.602 PMID:21328631

Lei, H.-T., Chou, T.-H., Su, C.-C., Bolla, J. R., Kumar, N., & Radhakrishnan, A. et al. (2014). Crystal Structure of the Open State of the *Neisseria gonorrhoeae* MtrE Outer Membrane Channel. *PLoS ONE*, *9*(6), e97475. doi:10.1371/journal.pone.0097475 PMID:24901251

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

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

Leo, A. J., & Hansch, C. (1999). Role of hydrophobic effects in mechanistic QSAR. *Perspectives in Drug Discovery and Design*, *17*(1), 1–25. doi:10.1023/A:1008762321231

Leonard, J. T., & Roy, K. (2006). On selection of training and test sets for the development of predictive QSAR models. *QSAR & Combinatorial Science*, *25*(3), 235–251. doi:10.1002/qsar.200510161

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

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

Levet, A., Bordes, C., Clément, Y., Mignon, P., Chermette, H., & Marote, P. et al. (2013). Quantitative structure-activity relationship to predict acute fish toxicity of organic solvents. *Chemosphere*, *93*, 1094–1103. PMID:23866172

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

Levitt, M., & Chothia, C. (1976). Structural patterns in globular proteins. *Nature*, *261*(5561), 552–558. doi:10.1038/261552a0 PMID:934293

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

Li, Z., Ning, C., Bin, Y., GuangFu, Y., Yun, L., & XinLing, Y. (2010). QSAR and 3D-QSAR studies of the diacyl-hydrazine derivatives containing furan rings based on the density functional theory. *SCIENCE CHINA Chemistry, 53*, 1322-1331.

Liao, S., Liang, T., Fang, S., Castenada, E., & Shao, T. C. (1973). Steroid structure and androgenic activity specificities involved in the receptor binding and nuclear retention of various androgens. *The Journal of Biological Chemistry*, *248*, 6154–6162. PMID:4353633

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

Lien, E. J., Hansch, C., & Anderson, S. M. (1968). Structure-activity correlations for antibacterial agents on gram-positive and gram-negative cells. *Journal of Medicinal Chemistry*, *11*(3), 430–441. doi:10.1021/jm00309a004 PMID:5656475

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.

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

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

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

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

Limbach, L. K., Li, Y., Grass, R. N., Brunner, T. J., Hintermann, M. A., & Muller, M. et al. (2005). Oxide nanoparticle uptake in human lung fibroblasts: Effects of particle size, agglomeration, and diffusion at low concentrations. *Environmental Science & Technology*, *39*(23), 9370–9376. doi:10.1021/es051043o PMID:16382966

Lin, D., & Xing, B. (2007). Phytotoxicity of nanoparticles: Inhibition of seed germination and root growth. *Environmental Pollution*, *150*(2), 243–250. doi:10.1016/j.envpol.2007.01.016 PMID:17374428

Lin, J., Michel, L. O., & Zhang, Q. (2002). CmeABC functions as a multidrug efflux system in Campylobacter jejuni. *Antimicrobial Agents and Chemotherapy*, *46*(7), 2124–2131. doi:10.1128/AAC.46.7.2124-2131.2002 PMID:12069964

Lin, S., Zhao, Y., Ji, Z., Ear, J., Chang, C. H., & Zhang, H. et al. (2013). Zebrafish high-throughput screening to study the impact of dissolvable metal oxide nanoparticles on the hatching enzyme, ZHE1. *Small*, *9*(9–10), 1776–1785. doi:10.1002/smll.201202128 PMID:23180726

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

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*, *23*(1-3), 4–25. doi:10.1016/S0169-409X(96)00423-1 PMID:11259830

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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, 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

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

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

Liu, R., Rallo, R., George, S., Ji, Z., Nair, S., Nel, A. E., & Cohen, Y. (2011). Classification NanoSAR development for cytotoxicity of metal oxide nanoparticles. *Small*, *7*(8), 1118–1126. doi:10.1002/smll.201002366 PMID:21456088

Liu, R., Rallo, R., Weissleder, R., Tassa, C., Shaw, S., & Cohen, Y. (2013). Nano-SAR development for bioactivity of nanoparticles with considerations of decision boundaries. *Small*, *9*(9-10), 1842–1852. doi:10.1002/smll.201201903 PMID:23423856

Liu, R., Zhang, H. Y., Ji, Z. X., Rallo, R., Xia, T., & Chang, C. H. et al. (2013). Development of structure-activity relationship for metal oxide nanoparticles. *Nanoscale*, *5*(12), 5644–5653. doi:10.1039/c3nr01533e PMID:23689214

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

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

Li, X.-Z., & Nikaido, H. (2009). Efflux-mediated drug resistance in bacteria. *Drugs*, *69*(12), 1555–1623. doi:10.2165/11317030-000000000-00000 PMID:19678712

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

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

Lomovskaya, O., Warren, M. S., Lee, A., Galazzo, J., Fronko, R., & Lee, M. et al. (2001). Identification and characterization of inhibitors of multidrug resistance efflux pumps in *Pseudomonas aeruginosa*: Novel agents for combination therapy. *Antimicrobial Agents and Chemotherapy*, *45*(1), 105–116. doi:10.1128/AAC.45.1.105-116.2001 PMID:11120952

Lomovskaya, O., & Watkins, W. (2001). Inhibition of efflux pumps as a novel approach to combat drug resistance in bacteria. *Journal of Molecular Microbiology and Biotechnology*, *3*(2), 225–236. PMID:11321578

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

Luan, F., Cordeiro, M. N. D. S., Alonso, N., Garcia-Mera, X., Caamano, O., & Romero-Duran, F. J. et al. (2013). TOPS-MODE model of multiplexing neuroprotective effects of drugs and experimental-theoretic study of new 1,3-rasagiline derivatives potentially useful in neurodegenerative diseases. *Bioorganic & Medicinal Chemistry*, *21*(7), 1870–1879. doi:10.1016/j.bmc.2013.01.035 PMID:23415089

Luan, F., Kleandrova, V. V., Gonzalez-Diaz, H., Ruso, J. M., Melo, A., Speck-Planche, A., & Cordeiro, M. N. (2014). Computer-aided nanotoxicology: Assessing cytotoxicity of nanoparticles under diverse experimental conditions by using a novel QSTR-perturbation approach. *Nanoscale*, *6*(18), 10623–10630. doi:10.1039/C4NR01285B PMID:25083742

Lubelski, J., Konings, W. N., & Driessen, A. J. (2007). Distribution and physiology of ABC-type transporters contributing to multidrug resistance in bacteria. *Microbiology and Molecular Biology Reviews*, *71*(3), 463–476. doi:10.1128/MMBR.00001-07 PMID:17804667

Lucic, B., & Trinajstic, N. (1999). Multivariate regression outperforms several robust architectures of neural networks in QSAR modeling. *Journal of Chemical Information and Computer Sciences*, *39*(1), 121–132. doi:10.1021/ci980090f

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

Lu, M., Symersky, J., Radchenko, M., Koide, A., Guo, Y., Nie, R., & Koide, S. (2013). Structures of a Na+-coupled, substrate-bound MATE multidrug transporter. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(6), 2099–2104. doi:10.1073/pnas.1219901110 PMID:23341609

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

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

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

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

Lyon, D. Y., Fortner, J. D., Sayes, C. M., Colvin, V. L., & Hughe, J. B. (2005). Bacterial cell association and antimicrobial activity of a C60 water suspension. *Environmental Toxicology and Chemistry*, *24*(11), 2757–2762. doi:10.1897/04-649R.1 PMID:16398110

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

Mackay, D. (1979). Finding fugacity feasible. *Environmental Science & Technology*, *13*(10), 1218–1223. doi:10.1021/es60158a003

Mackay, D., Di Guardo, A., Paterson, S., Kicsi, G., Cowan, C. E., & Kane, M. (1996). Assessment of chemical fate in the environment using evaluative, regional and local-scale models: Illustrative application to chlorobenzene and linear alkylbenzene sulfonates. *Environmental Toxicology and Chemistry*, *15*(9), 1638–1648. doi:10.1002/etc.5620150930

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Mackay, D., Shiu, W. Y., Ma, K.-C., & Lee, S. C. (2007). *Handbook of physical-chemical properties and environmental fate for organic chemicals*. Boca Raton, FL: Taylor & Francis.

MacLeod, M., Woodfine, D. G., Mackay, D., McKone, T., Bennett, D., & Maddalena, R. (2001). BETR North America: A regionally segmented multimedia contaminant fate model for North America. *Environmental Science and Pollution Research International*, *8*(3), 156–163. doi:10.1007/BF02987379 PMID:11505899

Maddalena, D. J. (1998). Applications of soft computing in drug design. *Expert Opinion on Therapeutic Patents*, *8*(3), 249–258. doi:10.1517/13543776.8.3.249

Mahamoud, A., Chevalier, J., Alibert-Franco, S., Kern, W. V., & Pagès, J.-M. (2007). Antibiotic efflux pumps in Gram-negative bacteria: The inhibitor response strategy. *The Journal of Antimicrobial Chemotherapy*, *59*(6), 1223–1229. doi:10.1093/jac/dkl493 PMID:17229832

Makarucha, A. J., Todorova, N., & Yarovsky, I. (2011). Nanomaterials in biological environment: A review of computer modelling studies. *European Biophysics Journal*, *40*(2), 103–115. doi:10.1007/s00249-010-0651-6 PMID:21153635

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

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

Malléa, M., Chevalier, J., Eyraud, A., & Pagès, J.-M. (2002). Inhibitors of antibiotic efflux pump in resistant *Enterobacter aerogenes* strains. *Biochemical and Biophysical Research Communications*, *293*(5), 1370–1373. doi:10.1016/S0006-291X(02)00404-7 PMID:12054665

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

Manallack, D. T., Tehan, B. G., Gancia, E., Hudson, B. D., Ford, M. G., & Livingstone, D. J. et al. (2003). A consensus neural network-based technique for discriminating soluble and poorly soluble compounds. *Journal of Chemical Information and Computer Sciences*, *43*(2), 674–679. doi:10.1021/ci0202741 PMID:12653537

Mannhold, R., Kubinyi, H., Folkers, G., Langer, T., & Hoffmann, R. D. (2006). *Pharmacophores and pharmacophore searches*. John Wiley & Sons.

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

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

Markham, P. N., & Neyfakh, A. A. (1996). Inhibition of the multidrug transporter NorA prevents emergence of norfloxacin resistance in Staphylococcus aureus. *Antimicrobial Agents and Chemotherapy*, *40*(11), 2673. PMID:8913490

Markham, P. N., & Neyfakh, A. A. (2001). Efflux-mediated drug resistance in Gram-positive bacteria. *Current Opinion in Microbiology*, *4*(5), 509–514. doi:10.1016/S1369-5274(00)00243-5 PMID:11587925

Marsalek, B., Jancula, D., Marsalkova, E., Mashlan, M., Safarova, K., Tucek, J., & Zboril, R. (2012). Multimodal action and selective toxicity of zerovalent iron nanoparticles against cyanobacteria. *Environmental Science & Technology*, *46*(4), 2316–2323. doi:10.1021/es2031483 PMID:22242974

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Martel, A. M., Leeson, P. A., & Castañer, J. (1997). BAY-12-8039. Fluoroquinolone antibacterial. *Drugs of the Future*, *22*(2), 109–113.

Martin, Y. C., & Lin, C. T. (1996). Article. In The practice of medicinal chemistry. London: Academic Press.

Martin, D., Maran, U., Sild, S., & Karelson, M. (2007). QSPR modeling solubility of polyaromatic hydrocarbons and fullerene in 1-octanol and n-heptane. *The Journal of Physical Chemistry B*, *111*(33), 9853–9857. doi:10.1021/jp071679x PMID:17661505

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

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

Martinez-Esparza, J., Oficialdegui, A. M., Perez-Silanes, S., Heras, B. B., Orus, L., & Palop, J. A. et al. (2001a). New 1-Aryl-3-(4-arylpiperazin-1-yl)propane derivatives, with dual action at 5-HT1A serotonin receptors and serotonin transporter, as a new class of antidepressants. *Journal of Medicinal Chemistry*, *44*(3), 418–428. doi:10.1021/jm001059j PMID:11462981

Martinez-Esparza, J., Perez-Silanes, S., Oficialdegui, A. M., Heras, B., Orus, L., & Villanueva, H. et al. (2001b). New 3-[4-(aryl)piperazin- 1-yl]-1-(benzo[b] thiophen-3-yl)propane derivatives with dual action at 5-HT1A serotonin receptors and serotonin transporter as a new class of antidepressants. *European Journal of Medicinal Chemistry*, *36*(1), 55–61. doi:10.1016/S0223-5234(00)01198-3 PMID:11231049

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

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

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

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.

Martin, Y. C. (2009). Let's not forget tautomers. *Journal of Computer-Aided Molecular Design*, *23*(10), 693–704. doi:10.1007/s10822-009-9303-2 PMID:19842045

Martin, Y. C., Kim, K.-H., & Lin, C. T. (1996). Article. *Advances in Quantitative Structure Property Relationships*, *1*, 1–52.

Matill, H. A. (1947). Antioxidants. *Annual Review of Biochemistry*, *16*(1), 177–192. doi:10.1146/annurev.bi.16.070147.001141 PMID:20259061

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

Matta, C. F. (2014). Modeling biophysical and biological properties from the characteristics of the molecular electron density, electron localization and delocalization matrices, and the electrostatic potential. *Journal of Computational Chemistry*, *35*(16), 1165–1198. doi:10.1002/jcc.23608 PMID:24777743

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

Ma, Y., He, X., Zhang, P., Zhang, Z., Guo, Z., & Tai, R. et al. (2011). Phytotoxicity and biotransformation of La_2O_3 nanoparticles in a terrestrial plant cucumber (Cucumis sativus). *Nanotoxicology*, *5*(4), 743–753. doi:10.3109/17435390.2010.545487 PMID:21261455

Ma, Y., Kuang, L., He, X., Bai, W., Ding, Y., & Zhang, Z. et al. (2010). Effects of rare earth oxide nanoparticles on root elongation of plants. *Chemosphere*, *78*(3), 273–279. doi:10.1016/j.chemosphere.2009.10.050 PMID:19897228

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

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

Maynard, A. T., Huang, M., Rice, W. G., & Covel, D. G. (1998). Reactivity of the HIV-1 nucleocapsid protein p7 zinc finger domains from the perspective of density-functional theory. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(20), 11578–11583. doi:10.1073/pnas.95.20.11578 PMID:9751708

Mazzariol, A., Tokue, Y., Kanegawa, T. M., Cornaglia, G., & Nikaido, H. (2000). High-level fluoroquinolone-resistant clinical isolates of Escherichia coli overproduce multidrug efflux protein AcrA. *Antimicrobial Agents and Chemotherapy*, *44*(12), 3441–3443. doi:10.1128/AAC.44.12.3441-3443.2000 PMID:11083655

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

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

McFarland, J. W. (1970). Parabolic relation between drug potency and hydrophobicity. *Journal of Medicinal Chemistry*, *13*(6), 1092–1196. doi:10.1021/jm00300a040 PMID:5479848

McKone, T. (1993). *CalTOX, a multimedia total-exposure model for hazardous-wastes sites part II: The dynamic multimedia transport and transformation model*. Retrieved from http://www.dtsc.ca.gov/AssessingRisk/upload/techman2.pdf

McMurry, L., Petrucci, R. E., & Levy, S. B. (1980). Active efflux of tetracycline encoded by four genetically different tetracycline resistance determinants in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, *77*(7), 3974–3977. doi:10.1073/pnas.77.7.3974 PMID:7001450

Mekenyan, O., & Bonchev, D. (1986). OASIS method for predicting biological activity of chemical compounds. *Acta Pharmaceutica Jugoslavica*, *36*, 225–237.

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Melagraki, G., Afantitis, A., Sarimveis, H., Igglessi-Markopouloua, O., & Supuran, C. T. (2006). QSAR study on para-substituted aromatic sulfonamides as carbonic anhydrase II inhibitors using topological information indices. *Bioorganic & Medicinal Chemistry*, *14*(4), 1108–1114. doi:10.1016/j.bmc.2005.09.038 PMID:16213737

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

Meng, J., Bai, H., Jia, M., Ma, X., Hou, Z., & Xue, X. et al. (2011). Restoration of antibiotic susceptibility in fluoroquinolone-resistant Escherichia coli by targeting acrB with antisense phosphorothioate oligonucleotide encapsulated in novel anion liposome. *The Journal of Antibiotics*, *65*(3), 129–134. doi:10.1038/ja.2011.125 PMID:22186597

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

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

Messina, G. A., Olsina, R. A., & Stege, P. W. (2013). Integration of nanomaterials in capillary and microchip electrophoresis as a flexible tool. In C. D. García, K. Y. Chumbimuni-Torres, & E. Carrilho (Eds.), *Capillary electrophoresis and microchip capillary electrophoresis, principles, applications, and limitations* (pp. 327–353). Hoboken, NJ: Wiley. doi:10.1002/9781118530009.ch19

METI. (1973). *Chemical substances control law ministry of economy, trade and industry*. METI.

Meyer, H. (1899). Zur theorie der alkoholnarkose: 1. Welche eigenschaft der anästhetica bedingt ihre narkotische wirkung? *ArchArchiv for Experimentelle Pathologie und Pharmakologie*, *42*, 109–118.

Meyer, H. H. (1899). On the theory of alcohol narcosis: First communication. Which property determines its narcotic effect? *Archives of Experimental Pathology and Pharmacology*, *42*, 109–118. doi:10.1007/BF01834479

Mezey, P. G. (1992). Shape-similarity measures for molecular bodies: A 3D topological approach to quantitative shape-activity relations. *Journal of Chemical Information and Computer Sciences*, *32*(6), 650–656. doi:10.1021/ci00010a011

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

Michalak, K., Wesolowska, O., Motohashi, N., Molnar, J., & Hendrich, A. (2006). Interactions of phenothiazines with lipid bilayer and their role in multidrug resistance reversal. *Current Drug Targets*, *7*(9), 1095–1105. doi:10.2174/138945006778226570 PMID:17017888

Michalet, S., Cartier, G., David, B., Mariotte, A.-M., Dijoux-franca, M.-G., & Kaatz, G. W. et al. (2007). *N*-Caffeoylphenalkylamide derivatives as bacterial efflux pump inhibitors. *Bioorganic & Medicinal Chemistry Letters*, *17*(6), 1755–1758. doi:10.1016/j.bmcl.2006.12.059 PMID:17275293

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

Miller, K. D.(2012). Inhibiting telomerase to reverse trastuzumab(T) resistance in HER2+ breast cancer. *Cancer Research*, *72*(24 suppl).

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. (1997). Aromatase inhibitors and breast cancer. *Cancer Treatment Reviews*, *23*(3), 171–187. doi:10.1016/S0305-7372(97)90037-2 PMID:9251721

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

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

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

Mirkin, B. (2005). *Clustering for data mining: A data recovery approach*. London: Chapman & Hall/CRC. doi:10.1201/9781420034912

Mishra, S. K., Mishra, V., Tripathi, P. N., & Khan, M. A. (2014). Quantum chemical descriptors based QSTR study of nitrobenzene derivatives against tetrahymena pyriformis. *Research Journal of Chemical Sciences*, *4*, 29–37.

Mitra, I., Saha, A., & Roy, K. (2010). Exploring quantitative structure–activity relationship studies of antioxidant phenolic compounds obtained from traditional Chinese medicinal plants. *Molecular Simulation*, *36*(13), 1067–1079. doi:10.1080/08927022.2010.503326

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Mitrano, D. M., Lesher, E. K., Bednar, A., Monserud, J., Higgins, C. P., & Ranville, J. F. (2012). Detecting nanoparticulate silver using single-particle inductively couple plasma-mass spectrometry. *Environmental Toxicology and Chemistry*, *31*(1), 115–121. doi:10.1002/etc.719 PMID:22012920

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

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

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

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

Mo, L.-Y., Liu, H.-Y., Yi, Z.-S., Li, Y.-H., & Dou, R.-N. (2011). QSTR study on the freshwater photobacteria toxicity of substituted benzenes. *Chinese Journal of Structural Chemistry*, *30*, 1403–1411.

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

Molnar, J., Hever, A., Fakla, I., Fischer, J., Ocsovski, I., & Aszalos, A. (1996). Inhibition of the transport function of membrane proteins by some substituted phenothiazines in E. coli and multidrug resistant tumor cells. *Anticancer Research*, *17*(1A), 481–486. PMID:9066699

Molyneaux, C.-A., Krugliak, M., Ginsburg, H., & Chibale, K. (2005). Arylpiperazines displaying preferential potency against chloroquine-resistant strains of the malaria parasite Plasmodium falciparum. *Biochemical Pharmacology*, *71*(1-2), 61–68. doi:10.1016/j.bcp.2005.10.023 PMID:16298348

Montgomery, D. C., & Peck, E. A. (1992). *Introduction to linear regression analysis*. New York: Wiley.

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

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

Morales Helguera, A., Perez Gonzalez, M., Cordeiro, M. N. D. S., & Cabrera Perez, M. A. (2008). Quantitative structure-carcinogenicity relationship for detecting structural alerts in nitroso compounds: Species, rat; sex, female; route of administration, gavage. *Chemical Research in Toxicology*, *21*(3), 633–642. doi:10.1021/tx700336n PMID:18293904

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

Moriguchi, I., Canada, Y., & Komatsu, K. (1976). van der Waals volume and the related parameters for hydrophobicity in structure-activity studies. *Chemical & Pharmaceutical Bulletin*, *24*(8), 1799–1806. doi:10.1248/cpb.24.1799

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., & 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

Mostrag, A., Puzyn, T., & Haranczyk, M. (2010). Modeling the overall persistence and environmental mobility of sulfur-containing polychlorinated organic compounds. *Environmental Science and Pollution Research International*, *17*(2), 470–477. doi:10.1007/s11356-009-0257-7 PMID:19937279

Moudgal, C. J., Lipscomb, J. C., & Bruce, R. M. (2000). Potential health effects of drinking water disinfection by-products using quantitative structure toxicity relationship. *Toxicology*, *147*(2), 109–131. doi:10.1016/S0300-483X(00)00188-8 PMID:10874158

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

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

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

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

Mullen, L. M. A., Duchowicz, P. R., & Castro, E. A. (2011). QSAR treatment on a new class of triphenylmethyl-containing compounds as potent anticancer agents. *Chemometrics and Intelligent Laboratory Systems*, *107*(2), 269–275. doi:10.1016/j.chemolab.2011.04.011

Mulliken, R. S. (1934). A new electroaffinity scale; together with data on valence states and on valence ionization potentials and electron affinities. *The Journal of Chemical Physics*, *2*(11), 782–793. doi:10.1063/1.1749394

Mulliken, R. S. (1935). Electronic structures of molecules XI. Electroaffinity, molecular orbitals and dipole moments. *The Journal of Chemical Physics*, *3*(9), 573–585. doi:10.1063/1.1749731

Mulliken, R. S. (1955). Electronic population analysis on LCAO-MO molecular wave functions. I. *The Journal of Chemical Physics*, *23*(10), 1833–1840. doi:10.1063/1.1740588

Munoz-Bellido, J., Munoz-Criado, S., & Garcìa-Rodrìguez, J. (2000). Antimicrobial activity of psychotropic drugs: Selective serotonin reuptake inhibitors. *International Journal of Antimicrobial Agents*, *14*(3), 177–180. doi:10.1016/S0924-8579(99)00154-5 PMID:10773485

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

Murakami, S., Nakashima, R., Yamashita, E., & Yamaguchi, A. (2002). Crystal structure of bacterial multidrug efflux transporter AcrB. *Nature*, *419*(6907), 587–593. doi:10.1038/nature01050 PMID:12374972

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Murray, J. S., & Politzer, P. (2003). The use of the molecular electrostatic potential in medicinal chemistry. In P. Carloni & F. Alber (Eds.), *Quantum medicinal chemistry*. Weinheim: Wiley-VCH.

Murray, J. S., & Politzer, P. A. (1994). General interaction properties function (GIPF): an approach to understanding and predicting molecular interactions. In P. Politzer & J. S. Murray (Eds.), *Quantitative treatments of solute/solvent interactions* (pp. 243–289). Amsterdam: Elsevier.

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

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

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

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

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 diarylalkylimidazole 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

Nakajoh, K., Shibata, E., Todoroki, T., Ohara, A., Nishizawa, K., & Nakamura, T. (2006). Measurement of temperature dependence for the vapor pressures of twenty six polychlorinated biphenyl congeners in commercial Kanechlor mixtures by the Knudsen effusion method. *Environmental Toxicology and Chemistry*, *25*(2), 327–336. doi:10.1897/05-215R.1 PMID:16519292

Nakashima, R., Sakurai, K., Yamasaki, S., Hayashi, K., Nagata, C., & Hoshino, K. et al. (2013). Structural basis for the inhibition of bacterial multidrug exporters. *Nature*, *500*(7460), 102–106. doi:10.1038/nature12300 PMID:23812586

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Nakayama, K., Kawato, H., Watanabe, J., Ohtsuka, M., Yoshida, K.-i., & Yokomizo, Y. et al. (2004). MexAB-OprM specific efflux pump inhibitors in *Pseudomonas aeruginosa*. Part 3: Optimization of potency in the pyridopyrimidine series through the application of a pharmacophore model. *Bioorganic & Medicinal Chemistry Letters*, *14*(2), 475–479. doi:10.1016/j.bmcl.2003.10.060 PMID:14698185

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

NanoMILE. (n.d.). Retrieved from http://www.nanomile.eu-vri.eu/

NanoPuzzles. (n.d.). Retrieved from http://www.nanopuzzles.eu/

NanoReg. (n.d.). Retrieved from http://www.nanoreg.eu/

NanoTox. (n.d.). Retrieved from http://www.nanotox.com/

Nantasenamat, C., Isarankura-Na-Ayudhya, C., Naenna, T., & Prachayasittikul, V. (2009). A practical overview of quantitative structure-activity relationship. *Excli Journal.*, *8*, 74–88.

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

Nargotra, A., Koul, S., Sharma, S., Khan, I., Kumar, A., & Thota, N. et al. (2009). Quantitative structure–activity relationship (QSAR) of aryl alkenyl amides/imines for bacterial efflux pump inhibitors. *European Journal of Medicinal Chemistry*, *44*(1), 229–238. doi:10.1016/j.ejmech.2008.02.015 PMID:18395941

Nargotra, A., Sharma, S., Koul, J. L., Sangwan, P. L., Khan, I. A., & Kumar, A. et al. (2009). Quantitative structure activity relationship (QSAR) of piperine analogs for bacterial NorA efflux pump inhibitors. *European Journal of Medicinal Chemistry*, *44*(10), 4128–4135. doi:10.1016/j.ejmech.2009.05.004 PMID:19523722

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

National Nanotechnology Initiative. (n.d.). *Collaborative central point of federal agencies in the USA*. Retrieved from www.nano.gov/about-nni

Nel, A., Xia, T., Madler, L., & Li, N. (2006). Toxic potential of materials at the nanolevel. *Science*, *311*(5761), 622–627. doi:10.1126/science.1114397 PMID:16456071

Nesmerak, K., Nemec, I., Sticha, M., Waisser, K., & Palat, K. (2005). Quantitative structure–property relationships of new benzoxazines and their electrooxidation as a model of metabolic degradation. *Electrochimica Acta*, *50*(6), 1431–1437. doi:10.1016/j.electacta.2004.08.031

Nesmerak, K., Toropov, A. A., & Toropova, A. (2014). SMILES-based quantitative structure–retention relationships for RP-HPLC of 1-phenyl-5-benzylsulfanyltetrazoles. *Structural Chemistry*, *25*(1), 311–317. doi:10.1007/s11224-013-0293-2

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Netzeva, T. I. (2005). Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships - The report and recommendations of ECVAM workshop 52. *ATLA Alternatives to Laboratory Animals, 33*, 155-173.

Netzeva, T. I., Worth, A., Aldenberg, T., Benigni, R., Cronin, M. T., Gramatica, P., … Yang, C. (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 Animal, 33*(2), 155-173.

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

Neu, H. C. (1992). The crisis in antibiotic resistance. *Science*, *257*(5073), 1064–1073. doi:10.1126/science.257.5073.1064 PMID:1509257

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

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

Nikaido, H. (2001). *Preventing drug access to targets: cell surface permeability barriers and active efflux in bacteria.* Paper presented at the Seminars in Cell & Developmental Biology. doi:10.1006/scdb.2000.0247

Nikaido, H. (1996). Multidrug efflux pumps of gram-negative bacteria. *Journal of Bacteriology*, *178*(20), 5853. PMID:8830678

Nikaido, H. (2003). Molecular basis of bacterial outer membrane permeability revisited. *Microbiology and Molecular Biology Reviews*, *67*(4), 593–656. doi:10.1128/MMBR.67.4.593-656.2003 PMID:14665678

Nikaido, H., Basina, M., Nguyen, V., & Rosenberg, E. Y. (1998). Multidrug efflux pump acrab of *salmonella typhimurium* excretes only those β-lactam antibiotics containing lipophilic side chains. *Journal of Bacteriology*, *180*(17), 4686–4692. PMID:9721312

Nikolic, S., Kovacevic, G., Milicevic, A., & Trinajstic, N. (2003). The Zagreb indices 30 years after. *Croatica Chemica Acta*, *76*(2), 113–124.

Nikolova-Jeliazkova, N., & Jaworska, J. (2005). An approach to determining applicability domain for QSAR group contribution models: An analysis of SRC KOWWIN. *Alternatives to Laboratory Animals*, *33*(5), 461–470. PMID:16268758

Nikolova, N., & Jaworska, J. (2004). Approaches to measure chemical similarity-a review. *QSAR & Combinatorial Science*, *22*(9-10), 1006–1026.

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

Norinder, U. (1998). Recent progress in CoMFA methodology and related techniques. *Perspectives in Drug Discovery and Design*, *12*, 25–39. doi:10.1023/A:1017057332537

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

Nowack, B., Rnaville, J. F., Diomond, S., Gallego-Urrea, J. A., Metcalfe, C., & Rose, J. et al.. (2012). Potential scenarios for nanomaterial release and subsequent alteration in the environment. *Environmental Toxicology and Chemistry*, *31*(1), 50–59. doi:10.1002/etc.726 PMID:22038832

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

Nunes, C. A., & Freitas, M. P. (2013). Introducing new dimensions in MIA-QSAR: A case for chemokine receptor inhibitors. *European Journal of Medicinal Chemistry*, *62*, 297–300. doi:10.1016/j.ejmech.2013.01.005 PMID:23357311

O'Brien, S. E., & Popelier, P. L. A. (1999). Quantum molecular similarity. Part 2: The relation between properties in BCP space and bond length. *Canadian Journal of Chemistry*, *77*(1), 28–36. doi:10.1139/v98-215

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

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

Oberdorster, G. (2010). Safety assessment for nanotechnology and nanomedicine: Concepts of nanotoxicology. *Journal of Internal Medicine*, *267*(1), 89–105. doi:10.1111/j.1365-2796.2009.02187.x PMID:20059646

Oberdörster, G., Maynard, A., Donaldson, K., Castranova, V., Fitzpatrick, J., & Ausman, K. et al. (2005). Principles for characterizing the potential human health effects from exposure to nanomaterials: Elements of a screening strategy. *Particle and Fibre Toxicology*, *6*(2), 8. doi:10.1186/1743-8977-2-8 PMID:16209704

Oberdörster, G., Stone, V., & Donaldson, K. (2007). Toxicology of nanoparticles: A historical perspective. *Nanotoxicology*, *1*(1), 2–25. doi:10.1080/17435390701314761

Oberg, T. (2005). A QSAR for the hydroxylradicalreaction rate constant: Validation, domain of application, and prediction. *Atmospheric Environment*, *39*(12), 2189–2200. doi:10.1016/j.atmosenv.2005.01.007

OECD. (2002). *Report of the OECD/UNEP workshop on the use of multimedia models for estimation overall environmental persistence and long-range Transport in the context of PBTs/POPs assessment*. OECD Environment, Health and Safety Publications.

OECD. (2004). *Guidance document on the use of multimedia models for estimatig overall environmental persistence and long-range transport*. Paris, France: Health and Safety Publications.

OECD. (2004). *Principles for the validation for regulatory purposes of (quantitative) structure-activity relationship models*. Paris, France: OECD.

OECD. (2004). *Principles for the validation of (Q)SARs*. Retrieved 31 March, 2014 from http://www.oecd.org/dataoecd/33/37/37849783.pdf

OECD. (2007) *Guidance document on the grouping of chemicals*. Organisation of Economic Cooperation and Development. Retrieved from http://www.oecd.org

OECD. (2007). Guidance document on the validation of (quantitative) structure activity relationship [(Q)SAR] models. no. 69. Organisation of Economic Cooperation and Development.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

OECD. (2007). *Quantitative structure-activity relationships project [(Q)SARs]*. Available online: http://search.oecd.org/officialdocuments/displaydocumentpdf/?doclanguage=en&cote=env/jm/mono(2007)2

Oethinger, M., & Levy, S. B. (2002). *CA patent no. 2378883*. C. I. P. Office.

Ojha, P. K., Mitra, I., Das, R. N., & Roy, K. (2011). Further exploring r^2_m metrics for validation of QSPR models. *Chemometrics and Intelligent Laboratory Systems*, *107*(1), 194–205. doi:10.1016/j.chemolab.2011.03.011

Ojha, P. K., & Roy, K. (2011). Comparative QSARs for antimalarial endochins: Importance of descriptor-thinning and noise reduction prior to feature selection. *Chemometrics and Intelligent Laboratory Systems*, *109*(2), 146–161. doi:10.1016/j.chemolab.2011.08.007

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

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

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

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

Olson, M. S., & Gurian, P. L. (2012). Risk assessment strategies as nanomaterials transition into commercial applications. *Journal of Nanoparticle Research*, *14*(4), 786. doi:10.1007/s11051-012-0786-8

Omote, H., Hiasa, M., Matsumoto, T., Otsuka, M., & Moriyama, Y. (2006). The MATE proteins as fundamental transporters of metabolic and xenobiotic organic cations. *Trends in Pharmacological Sciences*, *27*(11), 587–593. doi:10.1016/j.tips.2006.09.001 PMID:16996621

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

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

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

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

Overton, E. (1895). Über die osmotischen eigenschaften der lebenden pflanzen und tierzelle. *Vierteljahrsshriften Naturforschungen Ges Zurich*, *40*, 159–201.

Overton, E. (1901). *Studien Uber die Narkose*. Jena, Germany: Fischer.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Padmanabhan, J., Parthasarathi, R., Subramanian, V., & Chattaraj, P. K. (2006a). Theoretical study on the complete series of chloroanilines. *The Journal of Physical Chemistry A*, *110*(32), 9900–9907. doi:10.1021/jp061436p PMID:16898692

Padmanabhan, J., Parthasarathi, R., Subramanian, V., & Chattaraj, P. K. (2006b). QSPR models for polychlorinated biphenyls: *N*-Octanol/water partition coefficient. *Bioorganic & Medicinal Chemistry*, *14*(4), 1021–1028. doi:10.1016/j.bmc.2005.09.017 PMID:16214354

Padmanabhan, J., Parthasarathi, R., Subramanian, V., & Chattaraj, P. K. (2006c). Group philicity and electrophilicity as possible descriptors for modeling ecotoxicity applied to chlorophenols. *Chemical Research in Toxicology*, *19*(3), 356–364. doi:10.1021/tx050322m PMID:16544939

Padmanabhan, J., Parthasarathi, R., Subramanian, V., & Chattaraj, P. K. (2007). Using QSPR models to predict the enthalpy of vaporization of 209 polychlorinated biphenyl congeners. *QSAR & Combinatorial Science*, *26*(2), 227–237. doi:10.1002/qsar.200630027

Pagès, J.-M., Masi, M., & Barbe, J. (2005). Inhibitors of efflux pumps in gram-negative bacteria. *Trends in Molecular Medicine*, *11*(8), 382–389. doi:10.1016/j.molmed.2005.06.006 PMID:15996519

Palchaudhuri, R., & Hergenrother, P. J. (2008). Triphenylmethylamides (TPMAs): Structure-activity relationship of compounds that induce apoptosis in melanoma cells. *Bioorganic & Medicinal Chemistry Letters*, *18*(22), 5888–5891. doi:10.1016/j.bmcl.2008.07.128 PMID:18710803

Palchaudhuri, R., Nesterenko, V., & Hergenrother, P. J. (2008). The complex role of the triphenylmethyl motif in anticancer compounds. *Journal of the American Chemical Society*, *130*(31), 10274–10281. doi:10.1021/ja8020999 PMID:18611022

Pal, D. K., Purkayastha, S. K., Sengupta, C., & De, A. U. (1992). Quantitative structure-property relationships with TAU indices: Part I-Research octane numbers of alkane fuel molecules. *Indian Journal of Chemistry*, *31B*, 109–114.

Pal, D. K., Sengupta, C., & De, A. U. (1988). A new topochemical descriptor (TAU) in molecular connectivity concept: Part I – Aliphatic compounds. *Indian Journal of Chemistry*, *27B*, 734–739.

Pal, D. K., Sengupta, C., & De, A. U. (1989). Introduction of a novel topochemical index and exploitation of group connectivity concept to achieve predictability in QSAR and RDD. *Indian Journal of Chemistry*, *28B*, 261–267.

Pal, D. K., Sengupta, M., Sengupta, C., & De, A. U. (1990). QSAR with TAU (τ) indices: Part I – Polymethylene primary diamines as amebicidal agents. *Indian Journal of Chemistry*, *29B*, 451–454.

Pal, P., Mitra, I., & Roy, K. (2014). QSPR modeling of odor threshold of aliphatic alcohols using extended topochemical atom (ETA) indices. *Croatica Chemica Acta*, *87*(1), 29–37. doi:10.5562/cca2284

Pal, S., & Chandra, A. K. (1995). Some novel relationships of polarizability with dipole moments. *Journal of Physical Chemistry*, *99*(38), 13865–13867. doi:10.1021/j100038a017

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

Pan, D., Tseng, Y., & Hopfinger, A. J. (2003). Quantitative structure-based design: Formalism and application of receptor-dependent RD-4D-QSAR analysis to a set of glucose analogue inhibitors of glycogen phosphorylase. *Journal of Chemical Information and Computer Sciences*, *43*(5), 1591–1607. doi:10.1021/ci0340714 PMID:14502494

Pandith, A. H., & Islam, N. (2013). Comparative assessment of QSTR models based on density functional, hartree–fock, AM1, and PM3 methods for acute toxicity of aliphatic compounds toward Vibrio fischeri. *International Journal of Quantum Chemistry*, *113*(6), 830–839. doi:10.1002/qua.24133

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

Pan, S., & Chattaraj, P. K. (2013). Favorable direction in a chemical reaction through the maximum hardness principle. *Journal of the Mexican Chemical Society*, *57*, 23–24.

Pan, S., Solà, M., & Chattaraj, P. K. (2013). On the validity of the maximum hardness principle and the minimum electrophilicity principle during chemical reactions. *The Journal of Physical Chemistry A*, *117*(8), 1843–1852. doi:10.1021/jp312750n PMID:23373511

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

Pan, Y., Jiang, J., & Wang, Z. (2007). Quantitative structure–property relationship studies for predicting flash points of alkanes using group bond contribution method with back-propagation neural network. *Journal of Hazardous Materials*, *147*(1-2), 424–430. doi:10.1016/j.jhazmat.2007.01.025 PMID:17292543

Papa, E., van derWal, L., Arnot, J. A., & Gramatica, P. (2014). Metabolic biotransformation half-lives in fish: QSAR modeling and consensus analysis. *The Science of the Total Environment*, *470–471*, 1040–1046. doi:10.1016/j.scitotenv.2013.10.068 PMID:24239825

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

Park, Y., Lee, Y., Kim, H., Lee, Y., Yoon, Y., & Moon, B. et al. (2006). NMR data of flavone derivatives and their anti-oxidative activities. *Bulletin of the Korean Chemical Society*, *27*(10), 1537–1541. doi:10.5012/bkcs.2006.27.10.1537

Parr, R. G. (1922-1924). Szentpaly, Lv., & Liu, S. (1999). Electrophilicity index. *Journal of the American Chemical Society*, 121.

Parr, R. G., & Chattaraj, P. K. (1991). Principle of maximum hardness. *Journal of the American Chemical Society*, *113*(5), 1854–1855. doi:10.1021/ja00005a072

Parr, R. G., Donnelly, R. A., Levy, M., & Palke, W. E. (1978). Electronegativity: The density functional viewpoint. *The Journal of Chemical Physics*, *68*(8), 3801–3807. doi:10.1063/1.436185

Parr, R. G., & Pearson, R. G. (1983). Absolute hardness: Companion parameter to absolute electronegativity. *Journal of the American Chemical Society*, *105*(26), 7512–7516. doi:10.1021/ja00364a005

Parr, R. G., & Yang, W. (1984). Density functional approach to the frontier-electron theory of chemical reactivity. *Journal of the American Chemical Society*, *106*(14), 4049–4050. doi:10.1021/ja00326a036

Parr, R. G., & Yang, W. (1989). *Density functional theory of atoms and molecules*. New York: Oxford University Press.

Parr, R. G., & Yang, W. (1995). Density-functional theory of the electronic structure of molecules. *Annual Review of Physical Chemistry*, *46*(1), 701–728. doi:10.1146/annurev.pc.46.100195.003413 PMID:24341393

Parthasarathi, R., Elango, M., Subramanian, V., & Chattaraj, P. K. (2005). Variation of electrophilicity during molecular vibrations and internal rotations. *Theoretical Chemistry Accounts*, *113*(5), 257–266. doi:10.1007/s00214-005-0634-3

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Parthasarathi, R., Padmanabhan, J., Elango, M., Subramanian, V., & Chattaraj, P. K. (2004a). Intermolecular reactivity through the generalized philicity concept. *Chemical Physics Letters*, *394*(4-6), 225–230. doi:10.1016/j.cplett.2004.07.002

Parthasarathi, R., Subramanian, V., Roy, D. R., & Chattaraj, P. K. (2004b). Electrophilicity index as a possible descriptor of biological activity. *Bioorganic & Medicinal Chemistry*, *12*(21), 5533–5543. doi:10.1016/j.bmc.2004.08.013 PMID:15465330

Pasha, F. A., Srivastava, H. K., & Singh, P. P. (2005). QSAR study of estrogens with the help of PM3-based descriptors. *International Journal of Quantum Chemistry*, *104*(1), 87–100. doi:10.1002/qua.20569

Pasha, F. A., Srivastava, H. K., & Singh, P. P. (2005). Semiempirical QSAR study and ligand receptor interaction of estrogens. *Molecular Diversity*, *9*(1-3), 215–220. doi:10.1007/s11030-005-2711-x PMID:15789567

Pasqualini, J. R. (2004). The selective estrogen enzyme modulators in breast cancer: A review. *Biochimica et Biophysica Acta*, *1654*(2), 123–143. PMID:15172700

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

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

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

Pathak, R. K., & Gadre, S. R. (1990). Maximal and minimal characteristics of molecular electrostatic potentials. *The Journal of Chemical Physics*, *93*(3), 1770–1773. doi:10.1063/1.459703

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

Patil, G. S. (1991). Correlation of aqueous solubility and octanol-water partition coefficient based on molecular structure. *Chemosphere*, *22*(8), 723–738. doi:10.1016/0045-6535(91)90049-J

Pauling, L. (1932). The nature of the chemical bond. Iv. The energy of single bonds and the relative electronegativity of atoms. *Journal of the American Chemical Society*, *54*(9), 3570–3582. doi:10.1021/ja01348a011

Pauling, L. (1939). *The nature of the chemical bond*. Ithaca, NY: Cornell University Press.

Paulsen, I. T., Brown, M. H., & Skurray, R. A. (1996). Proton-dependent multidrug efflux systems. *Microbiological Reviews*, *60*(4), 575–608. PMID:8987357

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

Pearson, R. G. (1963). Hard and soft acids and bases. *Journal of the American Chemical Society*, *85*(22), 3533–3539. doi:10.1021/ja00905a001

Pearson, R. G. (1966). Acids and bases. *Science*, *151*(3707), 172–177. doi:10.1126/science.151.3707.172 PMID:17746330

Pearson, R. G. (1968a). Hard and soft acids and bases, HSAB, part 1: Fundamental principles. *Journal of Chemical Education*, *45*(9), 581–587. doi:10.1021/ed045p581

Pearson, R. G. (1986). Absolute electronegativity and hardness correlated with molecular orbital theory. *Proceedings of the National Academy of Sciences of the United States of America*, *83*(22), 8440–8441. doi:10.1073/pnas.83.22.8440 PMID:16578791

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Pearson, R. G. (1987). Recent advances in the concept of hard and soft acids and bases. *Journal of Chemical Education*, *64*(7), 561–567. doi:10.1021/ed064p561

Pearson, R. G. (1993). The principle of maximum hardness. *Accounts of Chemical Research*, *26*(5), 250–255. doi:10.1021/ar00029a004

Pearson, R. G. (1997). *Chemical hardness: Applications from molecules to solids*. Weinheim: Wiley-VCH. doi:10.1002/3527606173

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

Perkins, R., Fang, H., Tong, W., & Welsh, W. J. (2003). Quantitative structure-activity relationship methods: Perspectives on drug discovery and toxicology. *Environmental Toxicology and Chemistry*, *22*(8), 1666–1679. doi:10.1897/01-171 PMID:12924569

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

Phan, G., Benabdelhak, H., Lascombe, M.-B., Benas, P., Rety, S., & Picard, M. et al. (2010). Structural and dynamical insights into the opening mechanism of *p. aeruginosa* OprM channel. *Structure (London, England)*, *18*(4), 507–517. doi:10.1016/j.str.2010.01.018 PMID:20399187

Phosrithong, N., Samee, W., Nunthanavanit, P., & Ungwitayatorn, J. (2012). *In vitro* antioxidant activity study of novel chromone derivatives. *Chemical Biology & Drug Design*, *79*(6), 981–989. doi:10.1111/j.1747-0285.2012.01368.x PMID:22381130

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

Pieroni, M., Dimovska, M., Brincat, J. P., Sabatini, S., Carosati, E., & Massari, S. et al. (2010). From 6-amino-quinolone antibacterials to 6-amino-7-thiopyranopyridinylquinolone ethyl esters as inhibitors of *Staphylococcus aureus* multidrug efflux pumps. *Journal of Medicinal Chemistry*, *53*(11), 4466–4480. doi:10.1021/jm1003304 PMID:20446747

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

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

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

Platt, J. R. (1947). Influence of neighbor bonds on additive bond properties in paraffins. *Journal de Chimie Physique*, *15*, 419–420.

Plavšić, D., Nikolić, S., Trinajstić, N., & Mihalić, Z. (1993). On the Harary index for the characterization of chemical graphs. *Journal of Mathematical Chemistry*, *12*(1), 235–250. doi:10.1007/BF01164638

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

Poater, A., Saliner, A. G., Carbo-Dorca, R., Poater, J., Sola, M., Cavallo, L., & Worth, A. P. (2009). Modeling the structure-property relationships of nanoneedles: A journey toward nanomedicine. *Journal of Computational Chemistry*, *30*(2), 275–284. doi:10.1002/jcc.21041 PMID:18615420

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. (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

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

Politzer, P. (1987). A relationship between the charge capacity and the hardness of neutral atoms and groups. *The Journal of Chemical Physics*, *86*(2), 1072–1073. doi:10.1063/1.452296

Politzer, P. (1988). Computational approaches to the identification of suspect toxic molecules. *Toxicology Letters*, *43*(1-3), 257–276. doi:10.1016/0378-4274(88)90032-X PMID:3051523

Poole, K. (2000). Efflux-mediated resistance to fluoroquinolones in gram-negative bacteria. *Antimicrobial Agents and Chemotherapy*, *44*(9), 2233–2241. doi:10.1128/AAC.44.9.2233-2241.2000 PMID:10952561

Poole, K. (2005). Efflux-mediated antimicrobial resistance. *The Journal of Antimicrobial Chemotherapy*, *56*(1), 20–51. doi:10.1093/jac/dki171 PMID:15914491

Popat, K. C., Eltgroth, M., LaTempa, T. J., Grimes, C. A., & Desai, T. A. (2007). Titania nanotubes: A novel platform for drug-eluting coatings for medical implants? *Small*, *3*(11), 1878–1881. doi:10.1002/smll.200700412 PMID:17935080

Popelier, P. L. A. (1999). Quantum molecular similarity. 1. BCP space. *The Journal of Physical Chemistry A*, *103*(15), 2883–2890. doi:10.1021/jp984735q

Popelier, P. L. A., Smith, P. J., & Chaudry, U. A. (2004). Quantitative structure-activity relationships of mutagenic activity from quantum topological descriptors: Triazenes and halogenated hydroxyfuranones (mutagen-X) derivatives. *Journal of Computer-Aided Molecular Design*, *18*(11), 709–718. doi:10.1007/s10822-004-6815-7 PMID:15865063

Portney, N. G., & Ozkan, M. (2006). Nano-oncology: Drug delivery, imaging, and sensing. *Analytical and Bioanalytical Chemistry*, *384*(3), 620–630. doi:10.1007/s00216-005-0247-7 PMID:16440195

Porto, L. C., Souza, E. S., Junkes, B. S., Yunes, R. A., & Heinzen, V. E. F. (2008). Semi-empirical topological index: Development of QSPR/QSRR and optimization for alkylbenzenes. *Talanta*, *76*(2), 407–412. doi:10.1016/j.talanta.2008.03.023 PMID:18585298

Pos, K. M., & Diederichs, K. (2002). Purification, crystallization and preliminary diffraction studies of AcrB, an inner-membrane multi-drug efflux protein. *Acta Crystallographica. Section D, Biological Crystallography*, *58*(10), 1865–1867. doi:10.1107/S0907444902013963 PMID:12351840

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

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

Prado-Prado, F. J., Garcia, I., Garcia-Mera, X., & Gonzalez-Diaz, H. (2011). Entropy multi-target QSAR model for prediction of antiviral drug complex networks. *Chemometrics and Intelligent Laboratory Systems*, *107*(2), 227–233. doi:10.1016/j.chemolab.2011.02.003

Prado-Prado, F. J., Garcia-Mera, X., & Gonzalez-Diaz, H. (2010). Multi-target spectral moment QSAR versus ANN for antiparasitic drugs against different parasite species. *Bioorganic & Medicinal Chemistry*, *18*(6), 2225–2231. doi:10.1016/j.bmc.2010.01.068 PMID:20185316

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

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

Pramanik, S., & Roy, K. (2014). Modeling bioconcentration factor (BCF) using mechanistically interpretable descriptors computed from open source tool "PaDEL-Descriptor". *Environmental Science and Pollution Research International*, *21*(4), 2955–2965. doi:10.1007/s11356-013-2247-z PMID:24170502

Preparata, F. P., & Shamos, M. I. (1991). *Computational geometry: An introduction* (F. P. Preparata & M. I. Shamos, Eds.). New York: Springer-Verlag.

Prinn, R. G., Huang, J., Weiss, R. F., Cunnold, D. M., Fraser, P. J., & Simmonds, P. G. et al.. (2001). Evidence for substantial variations of atmospheric hydroxyl radicals in the past two decades. *Science*, *292*(5523), 1882–1888. doi:10.1126/science.1058673 PMID:11337586

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.

Pumbwe, L., & Piddock, L. J. (2002). Identification and molecular characterisation of CmeB, a *Campylobacter jejuni* multidrug efflux pump. *FEMS Microbiology Letters*, *206*(2), 185–189. doi:10.1111/j.1574-6968.2002.tb11007.x PMID:11814661

Puri, S., Chickos, J. S., & Welsh, W. J. (2001). Determination of vaporization enthalpies of polychlorinated biphenyls by correlation gas chromatography. *Analytical Chemistry*, *73*(7), 1480–1484. doi:10.1021/ac001246p PMID:11321297

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Putman, M., van Veen, H. W., & Konings, W. N. (2000). Molecular properties of bacterial multidrug transporters. *Microbiology and Molecular Biology Reviews*, *64*(4), 672–693. doi:10.1128/MMBR.64.4.672-693.2000 PMID:11104814

Putz, M. V., Ionascu, C., Putz, A.-M., & Ostafe, V. (2011). Alert-QSAR. Implications for electrophilic theory of chemical carcinogenesis. *International Journal of Molecular Sciences*, *12*(12), 5098–5134. doi:10.3390/ijms12085098 PMID:21954348

Puzyn, T. (2011). On the replacement of empirical parameters in multimedia mass balance models with QSPR data. *Journal of Hazardous Materials*, *192*(3), 970–977. doi:10.1016/j.jhazmat.2011.05.078 PMID:21741174

Puzyn, T., & Falandysz, J. (2005). Octanol/water partition coefficients of chloronaphthalenes. *Journal of Environmental Science and Health. Part A, Environmental Science and Engineering & Toxic and Hazardous Substance Control*, *40*(9), 1651–1663. doi:10.1081/ESE-200067976 PMID:16134358

Puzyn, T., & Falandysz, J. (2007). QSPR modeling of partition coefficients and Henry's law constants for 75 chloronaphthalene congeners by means of six chemometric approaches - A comparative study. *Journal of Physical and Chemical Reference Data*, *36*(1), 203–214. doi:10.1063/1.2432888

Puzyn, T., Falandysz, J., Jones, P. D., & Giesy, J. P. (2007). Quantitative structure - activity relationships for the prediction of relative in vitro potencies (REPs) for chloronaphthalenes. *Journal of Environmental Science and Health. Part A, Environmental Science and Engineering & Toxic and Hazardous Substance Control*, *42*(5), 573–590. doi:10.1080/10934520701244326 PMID:17454365

Puzyn, T., Leszczynska, D., & Leszczynski, J. (2009). Toward the development of "Nano-QSARs": Advances and challenges. *Small*, *5*(22), 2494–2509. doi:10.1002/smll.200900179 PMID:19787675

Puzyn, T., Leszczynska, D., & Leszczynski, J. (2009b). Quantitative structure-activity relationships (QSARs) in the European REACH system. Could these apporaches be applied to nanomaterials? In *Practical aspects of computational chemistry. Methods, concepts and applications.* Heidelberg, Germany: Springer. doi:10.1007/978-90-481-2687-3_9

Puzyn, T., Leszczyński, J., & Cronin, M. T. D. (2010). *Recent advances in QSAR studies: Methods and applications* (Vol. 8). Springer. doi:10.1007/978-1-4020-9783-6

Puzyn, T., Mostrag, A., Falandysz, J., Kholod, Y., & Leszczynski, J. (2009c). Predicting water solubility of congeners. Chloronaphthalenes - A case study. *Journal of Hazardous Materials*, *170*(2-3), 1014–1022. doi:10.1016/j.jhazmat.2009.05.079 PMID:19524360

Puzyn, T., Mostrąg, A., Suzuki, N., & Falandysz, J. (2008c). QSPR-based estimation of the atmospheric persistence for chloronaphthalene congeners. *Atmospheric Environment*, *42*(27), 6627–6636. doi:10.1016/j.atmosenv.2008.04.048

Puzyn, T., Mostrąg-Szlichtyng, A., Suzuki, N., & Haranczyk, M. (2009a). Metody chemometryczne w ocenie ryzyka. Ilościowe zależności pomiędzy strukturą chemiczną a właściwościami (QSPR) dla nowych rodzajów zanieczyszczeń chemicznych. In *Chemometria w nauce i praktyce*. Kraków: Wydawnictwo Instytutu Ekspertyz Sądowych.

Puzyn, T., Rasulev, B., Gajewicz, A., Hu, X., Dasari, T. P., & Michalkova, A. 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

Puzyn, T., Rostkowski, P., Swieczkowski, A., Jedrusiak, A., & Falandysz, J. (2006). Prediction of environmental partition coefficients and the Henry's law constants for 135 congeners of chlorodibenzothiophene. *Chemosphere*, *62*(11), 1817–1828. doi:10.1016/j.chemosphere.2005.07.055 PMID:16214201

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Puzyn, T., Suzuki, N., & Haranczyk, M. (2008b). How do the partitioning properties of polyhalogenated POPs change when chlorine is replaced by bromine? *Environmental Science & Technology*, *42*(14), 5189–5195. doi:10.1021/es8002348 PMID:18754368

Puzyn, T., Suzuki, N., Haranczyk, M., & Rak, J. (2008). Calculation of quantum-mechanical descriptors for QSPR at the DFT level. Is it necessary? *Journal of Chemical Information and Modeling*, *48*(6), 1174–1180. doi:10.1021/ci800021p PMID:18510372

Puzyn, T., Suzuki, N., & Sakurai, T. (2011). Estimating persistence of brominated and chlorinated organic pollutants in air, water, soil, and sediments with the QSPR-based classification scheme. *Molecular Diversity*, *15*(1), 173–188. doi:10.1007/s11030-010-9250-9 PMID:20386980

Quinn, D. M. (1987). Acetylcholinesterase: Enzyme structure, reaction dynamics, and virtual transition states. *Chemical Reviews*, *87*(5), 955–979. doi:10.1021/cr00081a005

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

Rai, M., Kon, K., Ingle, A., Duran, N., Galdiero, S., & Galdiero, M. (2014). Broad-spectrum bioactivities of silver nanoparticles: The emerging trends and future prospects. *Applied Microbiology and Biotechnology*, *98*(5), 1951–1961. doi:10.1007/s00253-013-5473-x PMID:24407450

Ramalho, T. C., Freitas, M. P., & da Cunha, E. F. F. (Eds.). (2012). *Chemoinformatics: Directions toward combating neglected diseases*. Dubai: Bentham.

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

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

Randić, M., & Razinger, M. (1996). On the characterization of three-dimensional molecular structure. In From chemical topology to three-dimensional. Plenum Press.

Randić, M. (1975). On characterization of molecular branching. *Journal of the American Chemical Society*, *97*(23), 6609–6615. doi:10.1021/ja00856a001

Randić, M. (1991). Generalized molecular descriptors. *Journal of Mathematical Chemistry*, *7*(1), 155–168. doi:10.1007/BF01200821

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

Randic, M., & Basak, S. C. (1999). Optimal molecular descriptors based on weighted path numbers. *Journal of Chemical Information and Computer Sciences*, *39*(2), 261–266. doi:10.1021/ci9800763

Randic, M., & Basak, S. C. (2001). New descriptor for structure-property and structure-activity correlations. *Journal of Chemical Information and Computer Sciences*, *41*(3), 650–656. doi:10.1021/ci000116e PMID:11410042

Randic, M., DeAlba, L. M., & Harris, F. E. (1998). Graphs with the same detour matrix. *Croatica Chemica Acta*, *71*(1), 53–68.

Randic, M., & Pompe, M. (2001). The variable connectivity index 1xf versus the traditional molecular descriptors: A comparative study of 1xf against descriptors of CODESSA. *Journal of Chemical Information and Computer Sciences*, *41*(3), 631–638. doi:10.1021/ci000119r PMID:11410040

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

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 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Rasulev, B. F., Saidkhodzhaev, A. I., Nazrullaev, S. S., Akhmedkhodzhaeva, K. S., Khushbaktova, Z. A., & Leszczynski, J. (2007). Molecular modelling and QSAR analysis of the estrogenic activity of terpenoids isolated from Ferula plants. *SAR and QSAR in Environmental Research*, *18*(7-8), 663–673. doi:10.1080/10629360701428631 PMID:18038366

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

Ravi Kumar, M. N. (2000). Nano and microparticles as controlled drug delivery devices. *Journal of Pharmacy & Pharmaceutical Sciences*, *3*(2), 234–258. PMID:10994037

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

Ray, S., & Roy, K. (2013). Modeling adsorption of organic compounds on activated carbon using ETA indices. *Chemical Engineering Science*, *104*, 427–438. doi:10.1016/j.ces.2013.09.018

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

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., 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

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

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

Rehn, B., Seiler, F., Rehn, S., Bruch, J., & Maier, M. (2003). Investigations on the inflammatory and genotoxic lung effects of two types of titanium dioxide: Untreated and surface treated. *Toxicology and Applied Pharmacology*, *189*(2), 84–95. doi:10.1016/S0041-008X(03)00092-9 PMID:12781626

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

Renau, T. E., Léger, R., Flamme, E. M., Sangalang, J., She, M. W., & Yen, R. et al. (1999). Inhibitors of efflux pumps in *Pseudomonas aeruginosa* potentiate the activity of the fluoroquinolone antibacterial levofloxacin. *Journal of Medicinal Chemistry*, *42*(24), 4928–4931. doi:10.1021/jm9904598 PMID:10585202

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Ren, B. (2003). Atom-type-based AI topological descriptors for quantitative structure–retention index correlations of aldehydes and ketones. *Chemometrics and Intelligent Laboratory Systems*, *66*(1), 29–39. doi:10.1016/S0169-7439(03)00004-2

Richardson, B. J. (1869). Physiological research on alcohols. *The Medical Times and Gazzette*, *2*, 703–706.

Richet, C. (1893). On the relationship between the toxicity and the physical properties of substances. *Comptes Rendus Société Biologie*, *5*, 775–776.

Richet, M. C. (1893). Sur le rapport entre la toxicité et les propriétés physiques des corps. *Comptes Rendus des Seances de la Société de Biologie et de ses Filiales*, *45*, 775–776.

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

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

Rohrbaugh, R. H., & Jurs, P. C. (1987). Descriptions of molecular shape applied in studies of structure/activity and structure/property relationships. *Analytica Chimica Acta*, *199*, 99–109. doi:10.1016/S0003-2670(00)82801-9

Rokhina, E. V., & Suri, R. P. S. (2012). Application of density functional theory (DFT) to study the properties and degradation of natural estrogen hormones with chemical oxidizers. *The Science of the Total Environment*, *417–418*, 280–290. doi:10.1016/j.scitotenv.2011.12.008 PMID:22245212

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

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

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

Rosenbusch, J. P. (2001). Stability of membrane proteins: Relevance for the selection of appropriate methods for high-resolution structure determinations. *Journal of Structural Biology*, *136*(2), 144–157. doi:10.1006/jsbi.2001.4431 PMID:11886216

Roy, D. R., Giri, S., & Chattaraj, P. K. (2009). Arsenic toxicity: An atom counting and electrophilicity-based protocol. *Molecular Diversity*, *13*(4), 551–556. doi:10.1007/s11030-009-9133-0 PMID:19326252

Roy, D. R., Pal, N., Mitra, A., Bultinck, P., Parthasarathi, R., Subramanian, V., & Chattaraj, P. K. (2007). An atom counting strategy towards analyzing the biological activity of sex hormones. *European Journal of Medicinal Chemistry*, *42*(11-12), 1365–1369. doi:10.1016/j.ejmech.2007.01.028 PMID:17416442

Roy, D. R., Parthasarathi, R., Padmanabhan, J., Sarkar, U., Subramanian, V., & Chattaraj, P. K. (2006a). Careful scrutiny of the philicity concept. *The Journal of Physical Chemistry A*, *110*(3), 1084–1093. doi:10.1021/jp053641v PMID:16420012

Roy, D. R., Parthasarathi, R., Subramanian, V., & Chattaraj, P. K. (2006b). An electrophilicity based analysis of toxicity of aromatic compounds towards Tetrahymena pyriformis. *QSAR & Combinatorial Science*, *25*(2), 114–122. doi:10.1002/qsar.200530146

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Roy, D. R., Sarkar, U., Chattaraj, P. K., Mitra, A., Padmanabhan, J., & Parthasarathi, R. et al. (2006c). Analyzing toxicity through electrophilicity. *Molecular Diversity*, *10*(2), 119–131. doi:10.1007/s11030-005-9009-x PMID:16763875

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, K., Chakraborty, P., Mitra, I., Ojha, P. K., Kar, S., & Das, R. N. (2013). Some case studies on application of "r_m^2" 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

Roy, K., & Das, R. N. (2010). QSTR with extended topochemical atom (ETA) indices. 14. QSAR modeling of toxicity of aromatic aldehydes to *Tetrahymena pyriformis*. *Journal of Hazardous Materials*, *183*(1−3), 913–922. doi:10.1016/j.jhazmat.2010.07.116 PMID:20739120

Roy, K., & Das, R. N. (2011a). On some novel extended topochemical atom (ETA) parameters for effective encoding of chemical information and modeling of fundamental physicochemical properties. *SAR and QSAR in Environmental Research*, *22*(5−6), 451–472. doi:10.1080/1062936X.2011.569900 PMID:21598192

Roy, K., & Das, R. N. (2011b). On extended topochemical atom (ETA) indices for QSPR studies. In E. A. Castro & A. K. Hagi (Eds.), *Advanced methods and applications in chemoinformatics: Research progress and new applications* (pp. 380–411). Hershey, PA: IGI Global.

Roy, K., & Das, R. N. (2012). QSTR with extended topochemical atom (ETA) indices. 15. Development of predictive models for toxicity of organic chemicals against fathead minnow using second generation ETA indices. *SAR and QSAR in Environmental Research*, *23*(1−2), 125–140. doi:10.1080/1062936X.2011.645872 PMID:22292780

Roy, K., & Das, R. N. (2013). QSTR with extended topochemical atom (ETA) indices. 16. Development of predictive classification and regression models for toxicity of ionic liquids towards *Daphnia magna*. *Journal of Hazardous Materials*, *254–255*, 166–178. doi:10.1016/j.jhazmat.2013.03.023 PMID:23608063

Roy, K., & Ghosh, G. (2003). Introduction of extended topochemical atom (ETA) Indices in the valence electron mobile (VEM) environment as tools for QSAR/QSPR studies. *Internet Electronic Journal of Molecular Design*, *2*(9), 599–620.

Roy, K., & Ghosh, G. (2004a). QSTR with extended topochemical atom indices. 2. Fish toxicity of substituted benzenes. *Journal of Chemical Information and Computer Sciences*, *44*(2), 559–567. doi:10.1021/ci0342066 PMID:15032536

Roy, K., & Ghosh, G. (2004b). QSTR with extended topochemical atom indices. 3. Toxicity of nitrobenzenes to *Tetrahymena pyriformis*. *QSAR & Combinatorial Science*, *23*(2–3), 99–108. doi:10.1002/qsar.200330864

Roy, K., & Ghosh, G. (2004c). 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., & Ghosh, G. (2005). QSTR with extended topochemical atom indices. Part 5. Modeling of the acute toxicity of phenylsulfonyl carboxylates to *Vibrio fischeri* using genetic function approximation. *Bioorganic & Medicinal Chemistry*, *13*(4), 1185–1194. doi:10.1016/j.bmc.2004.11.014 PMID:15670927

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Roy, K., & Ghosh, G. (2006a). QSTR with extended topochemical atom (ETA) indices. VI. Acute toxicity of benzene derivatives to tadpoles (*Rana japonica*). *Journal of Molecular Modeling*, *12*(3), 306–316. doi:10.1007/s00894-005-0033-7 PMID:16249936

Roy, K., & Ghosh, G. (2006b). QSTR with extended topochemical atom (ETA) indices. 8. QSAR for the inhibition of substituted phenols on germination rate of *Cucumis sativus* using chemometric tools. *QSAR & Combinatorial Science*, *25*(10), 846–859. doi:10.1002/qsar.200510211

Roy, K., & Ghosh, G. (2007). QSTR with extended topochemical atom (ETA) indices. 9. Comparative QSAR for the toxicity of diverse functional organic compounds to *Chlorella vulgaris* using chemometric tools. *Chemosphere*, *70*(1), 1–12. doi:10.1016/j.chemosphere.2007.07.037 PMID:17765287

Roy, K., & Ghosh, G. (2008). QSTR with extended topochemical atom indices. 10. Modeling of toxicity of organic chemicals to humans using different chemometric tools. *Chemical Biology & Drug Design*, *72*(5), 383–394. doi:10.1111/j.1747-0285.2008.00712.x PMID:19012574

Roy, K., & Ghosh, G. (2009a). QSTR with extended topochemical atom (ETA) indices. 11. Comparative QSAR of acute NSAID cytotoxicity in rat hepatocytes using chemometric tools. *Molecular Simulation*, *35*(8), 648–659. doi:10.1080/08927020902744664

Roy, K., & Ghosh, G. (2009b). QSTR with extended topochemical atom (ETA) indices. 12. QSAR for the toxicity of diverse aromatic compounds to *Tetrahymena pyriformis* using chemometric tools. *Chemosphere*, *77*(7), 999–1009. doi:10.1016/j.chemosphere.2009.07.072 PMID:19709717

Roy, K., & Ghosh, G. (2009c). QSTR with extended topochemical atom (ETA) Indices. 13. Modeling of hERG K^+ channel blocking activity of diverse functional drugs using different chemometric tools. *Molecular Simulation*, *35*(15), 1256–1268. doi:10.1080/08927020903015379

Roy, K., & Ghosh, G. (2010). Exploring QSARs with extended topochemical atom (ETA) indices for modeling chemical and drug toxicity. *Current Pharmaceutical Design*, *16*(24), 2625–2639. doi:10.2174/138161210792389270 PMID:20642426

Roy, K., & Kabir, H. (2012a). QSPR with extended topochemical atom (ETA) indices. Modeling of critical micelle concentration of non-ionic surfactants. *Chemical Engineering Science*, *73*, 86–98. doi:10.1016/j.ces.2012.01.005

Roy, K., & Kabir, H. (2012b). QSPR with extended topochemical atom (ETA) indices. 3. Modeling of critical micelle concentration of cationic surfactants. *Chemical Engineering Science*, *81*, 169–178. doi:10.1016/j.ces.2012.07.008

Roy, K., & Kabir, H. (2013). QSPR with extended topochemical atom (ETA) indices. Exploring effects of hydrophobicity, branching and electronic parameters on logCMC values of anionic surfactants. *Chemical Engineering Science*, *87*, 141–151. doi:10.1016/j.ces.2012.10.002

Roy, K., & Mitra, I. (2009). Advances in quantitative structure-activity relationship models of antioxidants. *Expert Opinion on Drug Discovery*, *4*(11), 1157–1175. doi:10.1517/17460440903307409 PMID:23480434

Roy, K., Mitra, I., Kar, S., Ojha, P. K., Das, R. N., & Kabir, H. (2012). Comparative studies on some metrics for external validation of QSPR models. *Journal of Chemical Information and Modeling*, *52*(2), 396–408. doi:10.1021/ci200520g PMID:22201416

Roy, K., Pal, D. K., De, A. U., & Sengupta, C. (2001). Comparative QSAR studies with molecular negentropy, molecular connectivity, STIMS and TAU indices: Part II- General anaesthetic activity of aliphatic hydrocarbons, halocarbons and ethers. *Indian Journal of Chemistry*, *40B*, 129–135.

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

Roy, K., & Saha, A. (2003a). Comparative QSPR studies with molecular connectivity, molecular negentropy and TAU indices part I: Molecular thermochemical properties of diverse functional acyclic compounds. *Journal of Molecular Modeling*, *9*(4), 259–270. doi:10.1007/s00894-003-0135-z PMID:12827454

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Roy, K., & Saha, A. (2003c). QSPR with TAU indices: Water solubility of diverse functional acyclic compounds. *Internet Electronic Journal of Molecular Design*, *2*, 475–491.

Roy, K., & Saha, A. (2005). QSPR with TAU indices: Molar refractivity of diverse functional acyclic compounds. *Indian Journal of Chemistry*, *44B*, 1693–1707.

Roy, K., & Sanyal, I. (2006). QSTR with extended topochemical atom indices. 7. QSAR of substituted benzenes to *Saccharomyces cerevisiae*. *QSAR & Combinatorial Science*, *25*(4), 359–371. doi:10.1002/qsar.200530172

Roy, K., Sanyal, I., & Ghosh, G. (2006a). QSPR of *n*-octanol/water partition coefficient of nonionic organic compounds using extended topochemical atom (ETA) indices. *QSAR & Combinatorial Science*, *26*(5), 629–646. doi:10.1002/qsar.200610112

Roy, K., Sanyal, I., & Roy, P. P. (2006b). QSPR of the bioconcentration factors of non-ionic organic compounds in fish using extended topochemical atom (ETA) indices. *SAR and QSAR in Environmental Research*, *17*(6), 563–582. doi:10.1080/10629360601033499 PMID:17162387

Roy, P. P., Paul, S., Mitra, I., & Roy, K. (2009). On two novel parameters for validation of predictive QSAR models. *Molecules (Basel, Switzerland)*, *14*(5), 1660–1701. doi:10.3390/molecules14051660 PMID:19471190

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

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

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

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

Ruiz, J. (2003). Mechanisms of resistance to quinolones: Target alterations, decreased accumulation and DNA gyrase protection. *The Journal of Antimicrobial Chemotherapy*, *51*(5), 1109–1117. doi:10.1093/jac/dkg222 PMID:12697644

Russom, C. L., Bradbury, S. P., Broderius, S. J., Hammermeister, D. E., & Drummond, R. A. (1997). Predicting modes of toxic action from chemical structure: Acute toxicity in the fathead minnow (*Pimephales promelas*). *Environmental Toxicology and Chemistry*, *16*(5), 948–967. doi:10.1897/1551-5028(1997)016<0948:PMOTAF>2.3.CO;2

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

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

Saartok, T., Dahlberg, E., & Gustafsson, J. A. (1984). Relative binding affinity of anabolic-androgenic steroids: Comparison of the binding to the androgen receptors in skeletal muscle and in prostate, as well as to sex hormone-binding globulin. *Endocrinology*, *114*(6), 2100–2106. doi:10.1210/endo-114-6-2100 PMID:6539197

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Sabatini, S., Gosetto, F., Serritella, S., Manfroni, G., Tabarrini, O., & Iraci, N. et al. (2012). Pyrazolo [4, 3-c] [1, 2] benzothiazines 5, 5-dioxide: A promising new class of *Staphylococcus aureus* NorA efflux pump inhibitors. *Journal of Medicinal Chemistry*, *55*(7), 3568–3572. doi:10.1021/jm201446h PMID:22432682

Sabatini, S., Kaatz, G. W., Rossolini, G. M., Brandini, D., & Fravolini, A. (2008). From phenothiazine to 3-phenyl-1, 4-benzothiazine derivatives as inhibitors of the *Staphylococcus aureus* NorA multidrug efflux pump. *Journal of Medicinal Chemistry*, *51*(14), 4321–4330. doi:10.1021/jm701623q PMID:18578473

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

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

Sadiq, I. M., Pakrashi, S., Chandrasekaran, N., & Mukherjee, A. (2011). Studies on toxicity of aluminum oxide (Al_2O_3) nanoparticles to microalgae species: *Scenedesmus sp.* and *Chlorella sp. Journal of Nanoparticle Research*, *13*(8), 3287–3299. doi:10.1007/s11051-011-0243-0

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

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

Saier, M. H. Jr. (1998). Molecular phylogeny as a basis for the classification of transport proteins from bacteria, archaea and eukarya. *Advances in Microbial Physiology*, *40*, 81–136. doi:10.1016/S0065-2911(08)60130-7 PMID:9889977

Saier, M. H. Jr, Beatty, J. T., Goffeau, A., Harley, K. T., Heijne, W., & Huang, S.-C. et al. (1999). The major facilitator superfamily. *Journal of Molecular Microbiology and Biotechnology*, *1*(2), 257–279. PMID:10943556

Sala, G., & Baldratti, G. (1957). Myotrophic (anabolic) activity of 4-substituted testosterone analogs.*Proceedings of the Society for Experimental Biology and Medicine New York*, *95*, 22-24. doi:10.3181/00379727-95-23105

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

Sanderson, R. T. (1952). Electronegativity I. Orbital electronegativity of neutral atoms. *Journal of Chemical Education*, *29*, 540–546.

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

Sayes, C. M., Fortner, J. D., Guo, W., Lyon, D., Boyd, A. M., & Ausman, K. D. et al. (2004). The differential cytotoxicity of water-soluble fullerenes. *Nano Letters*, *4*(10), 1881–1887. doi:10.1021/nl0489586

Sayes, C., & Ivanov, I. (2010). Comparative study of predictive computational models for nanoparticle-induced cytotoxicity. *Risk Analysis*, *30*(11), 1723–1734. doi:10.1111/j.1539-6924.2010.01438.x PMID:20561263

Schaafsma, G., Kroese, E. D., Tielemans, E. L. J. P., Van de Sandt, J. J. M., & Van Leeuwen, C. J. (2009). REACH, non-testing approaches and the urgent need for a change in mind set. *Regulatory Toxicology and Pharmacology*, *53*(1), 70–80. doi:10.1016/j.yrtph.2008.11.003 PMID:19061930

Scheringer, M., Wegmann, F., Fenner, K., & Hungerbuhler, K. (2000). Investigation of teh cold condensation of persistanet organic chemicals with global multimedia fate model. *Environmental Science & Technology*, *34*(9), 1842–1850. doi:10.1021/es991085a

Schindler, B. D., Jacinto, P., & Kaatz, G. W. (2013). Inhibition of drug efflux pumps in *Staphylococcus aureus*: Current status of potentiating existing antibiotics. *Future Microbiology*, *8*(4), 491–507. doi:10.2217/fmb.13.16 PMID:23534361

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

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

Schneider, L. S. (2000). The future of cholinergic replacement therapy in Alzheimer's disease. *Current Opinion in Investigational Drugs (London, England)*, *2*, 427–437.

Schowanek, D., Fox, K., Holt, M., Schroeder, F. R., Koch, V., & Cassani, G. et al. (2001). GREAT-ER: A new tool for management and risk assessment of chemicals in river basins. Contribution to GREAT-ER. *Water Science and Technology*, *43*, 179–185. PMID:11380178

Schowanek, D., & Webb, S. (2002). Exposure simulation for pharmaceuticals in European surface waters with GREAT-ER. *Toxicology Letters*, *131*(1-2), 39–50. doi:10.1016/S0378-4274(02)00064-4 PMID:11988357

Schrand, A. M., Dai, L. M., Schlager, J. J., & Hussain, S. M. (2012). Toxicity testing of nanomaterials. In New technologies for toxicity testing (pp. 58-75). New York: Landes Bioscience and Springer Science+Bussiness Media, LLC.

Schultz, T. W. (1999). Structure−toxicity relationships for benzenes evaluated with *Tetrahymena pyriformis. Chemical Research in Toxicology, 12*, 1262-1267.

Schultz, H. P. (1989). Topological organic chemistry. 1. Graph theory and topological indices of alkanes. *Journal of Chemical Information and Computer Sciences*, *29*(3), 227–228. doi:10.1021/ci00063a012

Schultz, T. W. (1997). Tetratox: Tetrahymena pyriformis population growth impairment endpointa surrogate for fish lethality. *Toxicology Mechanisms and Methods*, *7*(4), 289–309. doi:10.1080/105172397243079

Schultz, T. W., Cronin, M. T. D., & Netzeva, T. I. (2003a). The present status of QSAR in toxicology. *Journal of Molecular Structure THEOCHEM*, *622*(1-2), 23–38. doi:10.1016/S0166-1280(02)00615-2

Schultz, T. W., Cronin, M. T. D., Netzeva, T. I., & Aptula, A. O. (2002). Structure−toxicity relationships for aliphatic chemicals evaluated with *Tetrahymena pyriformis. Chemical Research in Toxicology*, *15*(12), 1602–1609. doi:10.1021/tx025589p PMID:12482243

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Schultz, T. W., Cronin, M. T. D., Walker, J. D., & Aptul, A. O. (2003b). Quantitative structure–activity relationships (QSARs) in toxicology: A historical perspective. *Journal of Molecular Structure THEOCHEM*, *622*(1-2), 1–22. doi:10.1016/S0166-1280(02)00614-0

Schultz, T. W., Netzeva, T. I., & Cronin, M. T. D. (2003c). Selection of data sets for QSARs: Analyses of Tetrahymena toxicity from aromatic compounds. *SAR and QSAR in Environmental Research*, *14*(1), 59–81. doi:10.1080/1062936021000058782 PMID:12688416

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

Schuur, J., Selzer, P., & Gasteiger, J. (1996). The coding of the three-dimensional structure of molecules by molecular transforms and its application to structure–spectra correlations and studies of biological activity. *Journal of Chemical Information and Computer Sciences*, *36*(2), 334–344. doi:10.1021/ci950164c

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

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

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

Seber, G. A. F. (1984). *Multivariate observations*. New York: John Wiley & Sons. doi:10.1002/9780470316641

Seeger, M. A., Schiefner, A., Eicher, T., Verrey, F., Diederichs, K., & Pos, K. M. (2006). Structural asymmetry of AcrB trimer suggests a peristaltic pump mechanism. *Science*, *313*(5791), 1295–1298. doi:10.1126/science.1131542 PMID:16946072

Seidell, A. (1912). A new bromine method for the determination of thymol, salicylates, and similar compounds. *Am. Chem. J.*, *47*, 508–526.

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

Selassie, C. D. (2003). Article. In D. J. Abraham (Ed.), Burger's medicinal chemistry and drug discovery (pp. 1–48). New York: John Wiley.

Selassie, C., & Verma, R. P. (2010). History of quantitative structure–activity relationships. In *Burger's medicinal chemistry and drug discovery*. Academic Press.

Selassie, C. D. (2003). *Burger's medicinal chemistry and drug discovery* (D. J. Abraham, Ed.). New York: John Wiley & Sons.

Selassie, C., Mekapati, S., & Verma, R. (2002). QSAR: Then and now. *Current Topics in Medicinal Chemistry*, *2*(12), 1357–1379. doi:10.2174/1568026023392823 PMID:12470285

Senior, S. A., Madbouly, M. D., & El massry, A.-M. (2011). QSTR of the toxicity of some organophosphorus compounds by using the quantum chemical and topological descriptors. *Chemosphere*, *85*(1), 7–12. doi:10.1016/j.chemosphere.2011.06.062 PMID:21757222

Sennhauser, G., Bukowska, M. A., Briand, C., & Grütter, M. G. (2009). Crystal structure of the multidrug exporter MexB from *Pseudomonas aeruginosa. Journal of Molecular Biology*, *389*(1), 134–145. doi:10.1016/j.jmb.2009.04.001 PMID:19361527

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

Serpone, N., Dondi, D., & Albini, A. (2007). Inorganic and organic UV filters: Their role and efficiency in sunscreens and suncare products. *Inorganica Chimica Acta*, *360*(3), 794–802. doi:10.1016/j.ica.2005.12.057

Serradel, M. N., Blancafort, P., & Castañer, J. (1983). DL-8280. *Drugs of the Future*, *8*(5), 395.

Seyoum, A., Asres, K., & El-Fiky, F. K. (2006). Structure–radical scavenging activity relationships of flavonoids. *Phytochemistry*, *67*(18), 2058–2070. doi:10.1016/j.phytochem.2006.07.002 PMID:16919302

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

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

Shaw, S. Y., Westly, E. C., Pittet, M. J., Subramanian, A., Schreiber, S. L., & Weissleder, R. (2008). Perturbational profiling of nanomaterial biologic activity. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(21), 7387–7392. doi:10.1073/pnas.0802878105 PMID:18492802

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

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

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

Sheridan, R., Feuston, R. P., Maiorov, V. N., & Kearsley, S. (2004). Similarity to molecules in the training set is a good discriminator for prediction accuracy in QSAR. *Journal of Chemical Information and Computer Sciences*, *44*(6), 1912–1928. doi:10.1021/ci049782w PMID:15554660

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

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

Shinohara, N., Matsumoto, K., Endoh, S., Maru, J., & Nakanishi, J. (2009). In vitro and in vivo genotoxicity tests on fullerene C60 nanoparticles. *Toxicology Letters*, *191*(2-3), 289–296. doi:10.1016/j.toxlet.2009.09.012 PMID:19772904

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

Shoichet, B. K. (2004). Virtual screening of chemical libraries. *Nature*, *432*(7019), 862–865. doi:10.1038/nature03197 PMID:15602552

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

Shvedova, A. A., & Kagan, V. E. (2010). The role of nanotoxicology in realizing the 'helping without harm' paradigm of nanomedicine: Lessons from studies of pulmonary effects of single-walled carbon nanotubes. *Journal of Internal Medicine*, *267*(1), 106–118. doi:10.1111/j.1365-2796.2009.02188.x PMID:20059647

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

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.

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

Silva, D. G., Freitas, M. P., da Cunha, E. F. F., Ramalho, T. C., & Nunes, C. A. (2012). Rational design of small modified peptides as ACE inhibitors. *Medicinal Chemistry Communications*, *3*(10), 1290–1293. doi:10.1039/c2md20214j

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

SIMCA-P 10.0. (2002). Retrieved from www.umetrics.com

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

Simon, F., Peterlik, H., Pfeiffer, R., Bernardi, J., & Kuzmany, H. (2007). Fullerene release from the inside of carbon nanotubes: A possible route toward drug delivery. *Chemical Physics Letters*, *445*(4-6), 288–292. doi:10.1016/j.cplett.2007.08.014

Simon-Manso, Y., & Fuentealba, P. (1998). On the density functional relationship between static dipole polarizability and global softness. *The Journal of Physical Chemistry A*, *102*(11), 2029–2032. doi:10.1021/jp972733k

Simon, Z. (1974). Specific interactions. Intermolecular forces, steric requirements, and molecular size. *Angewandte Chemie International Edition in English*, *13*(11), 719–727. doi:10.1002/anie.197407191

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

Singhal, K., Sahu, V. K., Singh, P., & Raj, P. (2014). DFT-based prediction of antifungal and insecticidal activities of perfluorophenyl antimony(III) and antimony(V) chlorides. *Medicinal Chemistry Research*, *23*(4), 1758–1767. doi:10.1007/s00044-013-0752-8

Singh, J., Dubey, R., & Atal, C. (1986). Piperine-mediated inhibition of glucuronidation activity in isolated epithelial cells of the guinea-pig small intestine: Evidence that piperine lowers the endogeneous UDP-glucuronic acid content. *The Journal of Pharmacology and Experimental Therapeutics*, *236*(2), 488–493. PMID:3080587

Singh, K. P., & Gupta, S. (2014). Nano-QSAR modeling for predicting biological activity of diverse nanomaterials. *Royal Society of Chemistry Advances*, *4*(26), 13215–13230.

Singh, P. P., Srivastava, H. K., & Pasha, F. A. (2004). DFT-based QSAR study of testosterone and its derivatives. *Bioorganic & Medicinal Chemistry*, *12*(1), 171–177. doi:10.1016/j.bmc.2003.11.002 PMID:14697782

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

Sippl, W., & Holtje, H. D. (2000). Structure-based 3D-QSAR—Merging the accuracy of structure-based alignments with the computational efficiency of ligand-based methods. *Journal of Molecular Structure THEOCHEM*, *503*(1-2), 31–50. doi:10.1016/S0166-1280(99)00361-9

Sissi, C., & Palumbo, M. (2010). In front of and behind the replication fork: Bacterial type IIA toposiomerases. *Cellular and Molecular Life Sciences*, *67*(12), 2001–2024. doi:10.1007/s00018-010-0299-5 PMID:20165898

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

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

Smellie, A. (2004). Accelerated K-means clustering in metric spaces. *Journal of Chemical Information and Computer Sciences*, *44*(6), 1929–1935. doi:10.1021/ci0499222 PMID:15554661

Smith, E. C., Williamson, E. M., Wareham, N., Kaatz, G. W., & Gibbons, S. (2007). Antibacterials and modulators of bacterial resistance from the immature cones of *Chamaecyparis lawsoniana. Phytochemistry*, *68*(2), 210–217. doi:10.1016/j.phytochem.2006.10.001 PMID:17109904

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

Smith, P. J., & Popelier, P. L. A. (2005). Quantum chemical topology (QCT) descriptors as substitutes for appropriate Hammett constants. *Organic & Biomolecular Chemistry*, *3*(18), 3399–3407. doi:10.1039/b507024d PMID:16132102

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

Song, J., Wang, Z., Findlater, A., Han, Z., Jiang, Z., & Chen, J. et al. (2013). Terpenoid mosquito repellents: A combined DFT and QSAR study. *Bioorganic & Medicinal Chemistry Letters*, *23*(5), 1245–1248. doi:10.1016/j.bmcl.2013.01.015 PMID:23375229

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Speck-Planche, A., Kleandrova, V. V., Luan, F., & Cordeiro, M. N. D. S. (2012). Predicting multiple ecotoxicological profiles in agrochemical fungicides: A multi-species chemoinformatic approach. *Ecotoxicology and Environment Safety, 80*, 308-313.

Speck-Planche, A., & Cordeiro, M. N. D. S. (2013). Simultaneous modeling of antimycobacterial activities and ADMET profiles: A chemoinformatic approach to medicinal chemistry. *Current Topics in Medicinal Chemistry*, *13*(14), 1656–1665. doi:10.2174/15680266113139990116 PMID:23889052

Speck-Planche, A., & Cordeiro, M. N. D. S. (2014). Simultaneous virtual prediction of anti-escherichia coli activities and ADMET profiles: A chemoinformatic complementary approach for high-throughput screening. *ACS Combinatorial Science*, *16*(2), 78–84. doi:10.1021/co400115s PMID:24383958

Speck-Planche, A., & Kleandrova, V. V. (2012). QSAR and molecular docking techniques for the discovery of potent monoamine oxidase B inhibitors: Computer-aided generation of new rasagiline bioisosteres. *Current Topics in Medicinal Chemistry*, *12*(16), 1734–1747. doi:10.2174/156802612803989282 PMID:23030609

Speck-Planche, A., Kleandrova, V. V., & Cordeiro, M. N. D. S. (2013a). New insights toward the discovery of antibacterial agents: Multi-tasking QSBER model for the simultaneous prediction of anti-tuberculosis activity and toxicological profiles of drugs. *European Journal of Pharmaceutical Sciences*, *48*(4-5), 812–818. doi:10.1016/j.ejps.2013.01.011 PMID:23376211

Speck-Planche, A., Kleandrova, V. V., Luan, F., & Cordeiro, M. N. D. S. (2013b). Multi-target inhibitors for proteins associated with Alzheimer: In silico discovery using fragment-based descriptors. *Current Alzheimer Research*, *10*(2), 117–124. doi:10.2174/1567205011310020001 PMID:22515494

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

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

SRC/SEMATECH Engineering Research Program. (n.d.). Retrieved from http://www.erc.arizona.edu/

Srivastava, H. K., Pasha, F. A., & Singh, P. P. (2005). Atomic softness-based QSAR study of testosterone. *International Journal of Quantum Chemistry*, *103*(3), 237–245. doi:10.1002/qua.20506

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

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

Stanforth, R. W., Kolossov, E., & Mirkin, B. (2007). A measure of domain of applicability for QSAR modeling based on intelligent K-means clustering. *QSAR & Combinatorial Science*, *26*(7), 837–844. doi:10.1002/qsar.200630086

Stankevich, I. V., Skovortsova, M. I., & Zefirov, N. S. (1995). On a quantum chemical interpretation of molecular connectivity indices for conjugated hydrocarbons. *Journal of Molecular Structure THEOCHEM*, *342*, 173–179. doi:10.1016/0166-1280(95)90111-6

Stanton, D. T., & Jurs, P. C. (1990). Development and use of charged partial surface area structural descriptors in computer-assisted quantitative structure–property relationship studies. *Analytical Chemistry*, *629*(21), 2323–2329. doi:10.1021/ac00220a013

Statsoft-Team. (2001). STATISTICA. Data analysis software system. Author.

Stavri, M., Piddock, L. J., & Gibbons, S. (2007). Bacterial efflux pump inhibitors from natural sources. *The Journal of Antimicrobial Chemotherapy*, *59*(6), 1247–1260. doi:10.1093/jac/dkl460 PMID:17145734

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

Stuper, A. J., Brugger, W. E., & Jurs, P. C. (1979). *Computer assisted studies of chemical structure and biological function*. New York: Wiley-Interscience.

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

Stupp, S., Donners, J., Li, L., & Mata, A. (2005). Expanding frontiers in biomaterials. *MRS Bulletin*, *30*(11), 864–873. doi:10.1557/mrs2005.276

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

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

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

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

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, 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

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, T. Y., Gottschalk, F., Hungerbuhler, K., & Nowack, B. (2014). Comprehensive probabilistic modelling of environmental emissions of engineered nanomaterials. *Environmental Pollution*, *185*, 69–76. doi:10.1016/j.envpol.2013.10.004 PMID:24220022

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Sushko, I., Novotarskyi, S., Körner, R., Pandey, A. K., Rupp, M., & Teetz, W. et al. (2011). Online chemical modeling environment (OCHEM): Web platform for data storage, model development and publishing of chemical information. *Journal of Computer-Aided Molecular Design*, *6*(6), 533–554. doi:10.1007/s10822-011-9440-2 PMID:21660515

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

Sylvester, J. J. (1878). Chemistry and algebra. *Nature*, *17*(432), 284–284. doi:10.1038/017284a0

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. (1956). *Steric effects in organic chemistry* (M. S. Newman, Ed.). New York: John Wiley and Sons.

Taft, R. W., Abboud, J.-L. M., Kamlet, M. J., & Abraham, M. H. (1985). Linear solvation energy relations. *Journal of Solution Chemistry*, *14*(3), 153–186. doi:10.1007/BF00647061

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

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

Talevi, A., Bellera, C. L., Ianni, M. D., Duchowicz, P. R., Bruno-Blanch, L. E., & Castro, E. A. (2012). An integrated drug development approach. Applying topological descriptors. *Current Computer-aided Drug Design*, *8*, 172–181. doi:10.2174/157340912801619076 PMID:22734704

Taniguchi, N. (1974). On the basic concept of 'nanotechnology'. Japan Society of Precision Engineering.

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

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

Tanwar, A., Pal, S., Roy, D. R., & Chattaraj, P. K. (2006). Minimum magnetizability principle. *The Journal of Chemical Physics*, *125*(5), 056101–056102. doi:10.1063/1.2227381 PMID:16942262

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

Taskinen, J., & Yliruusi, J. (2003). P rediction of physicochemical properties based on neural network modelling. *Advanced Drug Delivery Reviews*, *55*(9), 1163–1183. doi:10.1016/S0169-409X(03)00117-0 PMID:12954197

Tassa, C., Duffner, J. L., Lewis, T. A., Weissleder, R., Schreiber, S. L., Koehler, A. N., & Shaw, S. Y. (2010). Binding affinity and kinetic analysis of targeted small molecule-modified nanoparticles. *Bioconjugate Chemistry*, *21*(1), 14–19. doi:10.1021/bc900438a PMID:20028085

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Tawari, N., Lele, A., Khambete, M., & Degani, M. (2014). Mutagenicity prediction for nitroaromatic compounds using QSTR modeling. *International Journal of Pharmacy and Pharmaceutical Sciences*, *6*, 149–157.

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

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

Tegos, G. P., Haynes, M., Strouse, J. J., Khan, M. M. T., Bologa, C. G., Oprea, T. I., & Sklar, L. A. (2011). Microbial efflux pump inhibition: Tactics and strategies. *Current Pharmaceutical Design*, *17*(13), 1291. doi:10.2174/138161211795703726 PMID:21470111

Tenorio-Borroto, E., Garcia-Mera, X., Penuelas-Rivas, C. G., Vasquez-Chagoyan, J. C., Prado-Prado, F. J., Castanedo, N., & Gonzalez-Diaz, H. (2013). Entropy model for multiplex drug-target interaction endpoints of drug immunotoxicity. *Current Topics in Medicinal Chemistry*, *13*(14), 1636–1649. doi:10.2174/15680266113139990114 PMID:23889053

Tenorio-Borroto, E., Penuelas Rivas, C. G., Vasquez Chagoyan, J. C., Castanedo, N., Prado-Prado, F. J., Garcia-Mera, X., & Gonzalez-Diaz, H. (2012). ANN multiplexing model of drugs effect on macrophages; theoretical and flow cytometry study on the cytotoxicity of the anti-microbial drug G1 in spleen. *Bioorganic & Medicinal Chemistry*, *20*(20), 6181–6194. doi:10.1016/j.bmc.2012.07.020 PMID:22981917

Tenorio-Borroto, E., Penuelas-Rivas, C. G., Vasquez-Chagoyan, J. C., Castanedo, N., Prado-Prado, F. J., Garcia-Mera, X., & Gonzalez-Diaz, H. (2014). Model for high-throughput screening of drug immunotoxicity - Study of the anti-microbial G1 over peritoneal macrophages using flow cytometry. *European Journal of Medicinal Chemistry*, *72*, 206–220. doi:10.1016/j.ejmech.2013.08.035 PMID:24445280

Teófilo, R. F., Martins, J. P. A., & Ferreira, M. M. C. (2009). Sorting variables by using informative vectors as a strategy for feature selection in multivariate regression. *Journal of Chemometrics*, *23*(1), 32–48. doi:10.1002/cem.1192

Tetko, I. V. (2008). Associative neural network. *Methods in Molecular Biology (Clifton, N.J.)*, *458*, 185–202. doi:10.1007/978-1-60327-101-1_10 PMID:19065811

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

Tetko, I. V., Bruneau, P., Mewes, H. W., Rohrer, D. C., & Poda, G. I. (2006). Can we estimate the accuracy of ADME-Tox predictions? *Drug Discovery Today*, *11*(15-16), 700–707. doi:10.1016/j.drudis.2006.06.013 PMID:16846797

Tetko, I. V., Sushko, I., Pandey, A. K., Zhu, H., Tropsha, A., & Papa, E. et al. (2008). Critical assessment of QSAR models of environmental toxicity against Tetrahymena pyriformis: Focusing on applicability domain and overfitting by variable selection. *Journal of Chemical Information and Modeling*, *48*(9), 1733–1746. doi:10.1021/ci800151m PMID:18729318

Tetko, I. V., & Tanchuk, V. Y. (2002). Application of associative neural networks for prediction of lipophilicity in ALOGPS 2.1 program. *Journal of Chemical Information and Computer Sciences*, *42*(5), 1136–1145. doi:10.1021/ci025515j PMID:12377001

Tetko, I. V., Tanchuk, V. Yu., Kasheva, T. N., & Villa, A. E. P. (2001). Estimation of aqueous solubility of chemical compounds using E-state indices. *Journal of Chemical Information and Computer Sciences*, *41*(6), 1488–1493. doi:10.1021/ci000392t PMID:11749573

Thai, K., Ngo, T., Phan, T., Tran, T., Nguyen, N., Nguyen, T., & Le, M. (2014). Virtual screening for novel *Staphylococcus aureus* NorA efflux pump inhibitors from natural products. *Medicinal Chemistry (United Arab Emirates)*. doi: 10.2174/1573406410666140902110903

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

Thanabalu, T., Koronakis, E., Hughes, C., & Koronakis, V. (1998). Substrate-induced assembly of a contiguous channel for protein export from *E. coli*: Reversible bridging of an inner-membrane translocase to an outer membrane exit pore. *The EMBO Journal*, *17*(22), 6487–6496. doi:10.1093/emboj/17.22.6487 PMID:9822594

Thorarensen, A., Presley-Bodnar, A. L., Marotti, K. R., Boyle, T. P., Heckaman, C. L., & Bohanon, M. J. et al. (2001). 3-Arylpiperidines as potentiators of existing antibacterial agents. *Bioorganic & Medicinal Chemistry Letters*, *11*(14), 1903–1906. doi:10.1016/S0960-894X(01)00330-4 PMID:11459657

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

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

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

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

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. (2009). *Molecular descriptors for chemoinformatics*. Weinheim: WILEY-VCH Verlag GmbH & Co. KGaA.

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

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. (2000). *Handbook of molecular descriptors* (Vol. 11). Weinheim: WILEY-VCH Verlag GmbH. doi:10.1002/9783527613106

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

Todeschini, R., Lasagni, M., & Marengo, E. (1994). New molecular descriptors for 2D- and 3Dstructures. Theory. *Journal of Chemometrics*, *8*(4), 263–273. doi:10.1002/cem.1180080405

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

Tong, W., Hong, H., Fang, H., Xie, Q., & Perkins, R. (2003). Decision forest: Combining the predictions of multiple independent decision tree models. *Journal of Chemical Information and Computer Sciences*, *43*(2), 525–531. doi:10.1021/ci020058s PMID:12653517

Tong, W., Hong, H., Xie, Q., Xie, L., Fang, H., & Perkins, R. (2004). Assessing QSAR limitations - A regulatory perspective. *Current Computer-aided Drug Design*, *1*(2), 195–205. doi:10.2174/1573409053585663

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Tong, W., Xie, Q., Hong, H., Shi, L., Fang, H., & Perkins, R. (2004). Assessment of prediction confidence and domain extrapolation of two structure–activity relationship models for predicting estrogen receptor binding activity. *Environmental Health Perspectives*, *112*(12), 1249–1254. doi:10.1289/ehp.7125 PMID:15345371

TOPKAT OPS. (2000). *U.S. Patent 6, 036, 349, March 14, 2000*. Washington, DC: US Patent Office.

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

Topliss, J. G. (1983). *Quantitative structure-activity relationships of drugs*. New York, NY: Academic Press.

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

Toropov, A. A., Leszczynska, D., & Leszczynski, J. (2007). Predicting water solubility and octanolwater partition coefficient for carbon nanatubes based on the chiral vector. *Computational Biology and Chemistry*, *31*(2), 127–128. doi:10.1016/j.compbiolchem.2007.02.002 PMID:17363328

Toropov, A. A., & Leszczynski, J. (2007). A new approach to the characterization of nanomaterials: Predicting young's modulus by correlation weighting of nanomaterials codes. *Chemical Physics Letters*, *433*(1-3), 125–129. doi:10.1016/j.cplett.2006.11.010

Toropov, A. A., Nesmerak, K., Raska, I. Jr, Waisser, K., & Palat, K. (2006). QSPR modeling of the half-wave potentials of benzoxazines by optimal descriptors calculated with the SMILES. *Computational Biology and Chemistry*, *30*(6), 434–437. doi:10.1016/j.compbiolchem.2006.09.003 PMID:17092778

Toropov, A. A., Rasulev, B. F., & Leszczynski, J. (2008). QSAR modeling of acute toxicity by balance of correlations. *Bioorganic & Medicinal Chemistry*, *16*(11), 5999–6008. doi:10.1016/j.bmc.2008.04.055 PMID:18482841

Toropov, A. A., & Toropova, A. P. (2001). Prediction of heteroaromatic amine mutagenicity by means of correlation weighting of atomic orbital graphs of local invariants. *Journal of Molecular Structure THEOCHEM*, *538*(1-3), 287–293. doi:10.1016/S0166-1280(00)00713-2

Toropov, A. A., & Toropova, A. P. (2002). QSAR modeling of toxicity on optimization of correlation weights of Morgan extended connectivity. *Journal of Molecular Structure THEOCHEM*, *578*(1-3), 129–134. doi:10.1016/S0166-1280(01)00695-9

Toropov, A. A., & Toropova, A. P. (2014). Optimal descriptor as a translator of eclectic data into endpoint prediction: Mutagenicity of fullerene as a mathematical function of conditions. *Chemosphere*, *104*, 262–264. doi:10.1016/j.chemosphere.2013.10.079 PMID:24246220

Toropov, A. A., Toropova, A. P., & Benfenati, E. (2010b). SMILES-based optimal descriptors: QSAR modeling of carcinogenicity by balance of correlations with ideal slopes. *European Journal of Medicinal Chemistry*, *45*(9), 3581–3587. doi:10.1016/j.ejmech.2010.05.002 PMID:20570021

Toropov, A. A., Toropova, A. P., Benfenati, E., Gini, G., & Fanelli, R. (2013b). The definition of the molecular structure for potential anti-malaria agents by the Monte Carlo method. *Structural Chemistry*, *24*(4), 1369–1381. doi:10.1007/s11224-012-0180-2

Toropov, A. A., Toropova, A. P., Benfenati, E., Gini, G., Leszczynska, D., & Leszczynski, J. (2010c). QSAR analysis of 1,4-dihydro-4-oxo-1-(2-thiazolyl)-1,8-naphthyridines exhibiting anticancer activity by optimal SMILES-based descriptors. *Journal of Mathematical Chemistry*, *47*(2), 647–666. doi:10.1007/s10910-009-9593-6

Toropov, A. A., Toropova, A. P., Benfenati, E., Gini, G., Leszczynska, D., & Leszczynski, J. (2013a). CORAL: QSPR model of water solubility based on local and global SMILES attributes. *Chemosphere*, *90*(2), 877–880. doi:10.1016/j.chemosphere.2012.07.035 PMID:22921649

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Toropov, A. A., Toropova, A. P., Benfenati, E., Gini, G., Puzyn, T., Leszczynska, D., & Leszczynski, J. (2012). Novel application of the CORAL software to model cytotoxicity of metal oxide nanoparticles to bacteria Escherichia coli. *Chemosphere*, *89*(9), 1098–1102. doi:10.1016/j.chemosphere.2012.05.077 PMID:22704203

Toropov, A. A., Toropova, A. P., Benfenati, E., Leszczynska, D., & Leszczynski, J. (2010d). SMILES-based optimal descriptors: QSAR analysis of fullerene-based HIV-1 PR inhibitors by means of balance of correlations. *Journal of Computational Chemistry*, *31*, 381–392. PMID:19479738

Toropov, A. A., Toropova, A. P., Puzyn, T., Benfenati, E., Gini, G., Leszczynska, D., & Leszczynski, J. (2013). QSAR as a random event: Modeling of nanoparticles uptake in PaCa2 cancer cells. *Chemosphere*, *92*(1), 31–37. doi:10.1016/j.chemosphere.2013.03.012 PMID:23566368

Toropov, A. A., Toropova, A. P., Raska, I. Jr, & Benfenati, E. (2010a). QSPR modeling of octanol/water partition coefficient of antineoplastic agents by balance of correlation. *European Journal of Medicinal Chemistry*, *45*(4), 1639–1647. doi:10.1016/j.ejmech.2009.12.043 PMID:20106561

Toropov, A. A., Toropova, A. P., Raska, I. Jr, Benfenati, E., & Gini, G. (2012b). QSAR modeling of endpoints for peptides which is based on representation of the molecular structure by a sequence of amino acids. *Structural Chemistry*, *23*(6), 1891–1904. doi:10.1007/s11224-012-9995-0

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

Toropov, A. A., Toropova, A. P., Rasulev, B. F., Benfenati, E., Gini, G., Leszczynska, D., & Leszczynski, J. (2012a). Coral: QSPR modeling of rate constants of reactions between organic aromatic pollutants and hydroxyl radical. *Journal of Computational Chemistry*, *33*(23), 1902–1906. doi:10.1002/jcc.23022 PMID:22641453

Toropova, A. P., & Toropov, A. A. (2013). Optimal descriptor as a translator of eclectic information into the prediction of membrane damage by means of various TiO_2 nanoparticles. *Chemosphere*, *93*(10), 2650–2655. doi:10.1016/j.chemosphere.2013.09.089 PMID:24161577

Toropova, A. P., & Toropov, A. A. (2014). CORAL software: Prediction of carcinogenicity of drugs by means of the Monte Carlo method. *European Journal of Pharmaceutical Sciences*, *52*, 21–25. doi:10.1016/j.ejps.2013.10.005 PMID:24514451

Toropova, A. P., Toropov, A. A., Benfenati, E., Gini, G., Leszczynska, D., & Leszczynski, J. (2011a). CORAL: Quantitative structure-activity relationship models for estimating toxicity of organic compounds in rats. *Journal of Computational Chemistry*, *32*(12), 2727–2733. doi:10.1002/jcc.21848 PMID:21656789

Toropova, A. P., Toropov, A. A., Benfenati, E., Gini, G., Leszczynska, D., & Leszczynski, J. (2011b). CORAL: QSPR models for solubility of [C60] and [C70] fullerene derivatives. *Molecular Diversity*, *15*(1), 249–256. doi:10.1007/s11030-010-9245-6 PMID:20349134

Toropova, A. P., Toropov, A. A., Benfenati, E., Gini, G., Leszczynska, D., & Leszczynski, J. (2012b). CORAL: Models of toxicity of binary mixtures. *Chemometrics and Intelligent Laboratory Systems*, *119*, 39–43. doi:10.1016/j.chemolab.2012.10.001

Toropova, A. P., Toropov, A. A., Benfenati, E., Leszczynska, D., & Leszczynski, J. (2010). QSAR modeling of measured binding affinity for fullerene-based HIV-1 PR inhibitors by CORAL. *Journal of Mathematical Chemistry*, *48*(4), 959–987. doi:10.1007/s10910-010-9719-x

Toropova, A. P., Toropov, A. A., Martyanov, S. E., Benfenati, E., Gini, G., Leszczynska, D., & Leszczynski, J. (2012a). CORAL: QSAR modeling of toxicity of organic chemicals towards Daphnia magna. *Chemometrics and Intelligent Laboratory Systems*, *110*(1), 177–181. doi:10.1016/j.chemolab.2011.10.005

Tran, T.-D., Park, H., Kim, H. P., Ecker, G. F., & Thai, K.-M. (2009). Inhibitory activity of prostaglandin E2 production by the synthetic 2′-hydroxychalcone analogues: Synthesis and SAR study. *Bioorganic & Medicinal Chemistry Letters*, *19*(6), 1650–1653. doi:10.1016/j.bmcl.2009.02.001 PMID:19233646

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Traube, J. (1904). Theorie der osmose and narkose. *Archiv für die gesamte Physiologie des Menschen und der Tiere, 105*(11−12), 541−558.

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

Tripathi, P. N., & Mishra, V. (2014). Toxicological study of nitrobenzene derivatives against tetrahymena pyriformis using topological parameters. *Journal of Chemical and Pharmaceutical Research*, *6*, 848–854.

Trohalaki, S., & Pachter, R. (2003). Quantum descriptors for predictive toxicology of halogenated aliphatic hydrocarbons. *SAR and QSAR in Environmental Research*, *14*(2), 131–143. doi:10.1080/1062936031000073153 PMID:12747571

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

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

Troshin, P. A., Hoppe, H., Renz, J., Egginger, M., Mayorova, J. Yu., & Goryachev, A. E. et al. (2009). Material solubility-photovoltaic performance relationship in the design of novel fullerene derivatives for bulk heterojunction solar cells. *Advanced Functional Materials*, *19*(5), 779–788. doi:10.1002/adfm.200801189

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

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

Tseng, T.-T., Gratwick, K. S., Kollman, J., Park, D., Nies, D. H., Goffeau, A., & Saier, M. H. Jr. (1999). The RND permease superfamily: An ancient, ubiquitous and diverse family that includes human disease and development proteins. *Journal of Molecular Microbiology and Biotechnology*, *1*(1), 107–125. PMID:10941792

Turabekova, M. A., Rasulev, B. F., Dzhakhangirov, F. N., Toropov, A. A., Leszczynska, D., & Leszczynski, J. (2014). *Aconitum* and *Delphinium* diterpenoid alkaloids of local anesthetic activity. Comparative QSAR analysis based on GA-MLRA/PLS and optimal descriptors approach. *Journal of Environmental Science and Health*, *32*(Part C), 213–238. doi:10.1080/10590501.2014.938886 PMID:25226219

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

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

UN-ECE. (1998). *Protocol to the 1979 convention on long range transboundary air pollution on persistent organic pollutants and executive body decision 1998/2 on information to be submitted and the procedure for adding substances to annexes I, II or III to the protocol on persistent organic pollutants. Document. ECE/EB.AIR/60*. New York: United Nations.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

UNEP. (2001). *Final act of the conference of plenipotentiaries on the Stockholm convention on persistent organic pollutants. Document. UNEP/POPS/CONF/4*. Stockholm: United Nations Environment Programme.

USEPA. (1999). *Toxic substances control act*. The United States Environmental Protection Agency.

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

Vamecq, J., Lambert, D., Poupaert, J. H., Masereel, B., & Stables, J. P. (1998). Anticonvulsant activity and interactions with neuronal voltage-dependent sodium channel of analogues of ameltolide. *Journal of Medicinal Chemistry*, *41*(18), 3307–3313. doi:10.1021/jm9608772 PMID:9719582

Van Bambeke, F., Balzi, E., & Tulkens, P. M. (2000). Antibiotic efflux pumps. *Biochemical Pharmacology*, *60*(4), 457–470. doi:10.1016/S0006-2952(00)00291-4 PMID:10874120

Van Bambeke, F., & Lee, V. J. (2006). Inhibitors of bacterial efflux pumps as adjuvants in antibiotic treatments and diagnostic tools for detection of resistance by efflux. *Recent Patents on Anti-infective Drug Discovery*, *1*(2), 157–175. doi:10.2174/157489106777452692 PMID:18221142

van de Waterbeemd, H. (1995). *Chemometrics methods in molecular design*. Weinheim: VCH Publishers. doi:10.1002/9783527615452

van der Meent, D. (1993). *SimpleBox: A generic multimedia fate evaluation model*. Institute of Public Health and the Environment.

van Drie, J. H. (2007). Monty Kier and the origin of the pharmacophore concept. *Internet Electronic Journal of Molecular Design*, *6*(9), 271–279.

van Vleck, J. (1932). *The theory of electric and magnetic susceptibilities*. Oxford, UK: Oxford University Press.

van Wezel, A. P., & Jager, T. (2002). Comparison of two screening level risk assessment approaches for six disinfectants and pharmaceuticals. *Chemosphere*, *47*(10), 1113–1128. doi:10.1016/S0045-6535(02)00048-6 PMID:12137045

van Wijk, D., Chenier, R., Henry, T., Hernando, M. D., & Schulte, C. (2009). Integrated approach to PBT and POP prioritization and risk assessment. *Integrated Environmental Assessment and Management*, *5*(4), 697–711. doi:10.1897/IEAM_2009-034.1 PMID:19552502

Vargiu, A. V., & Nikaido, H. (2012). Multidrug binding properties of the AcrB efflux pump characterized by molecular dynamics simulations. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(50), 20637–20642. doi:10.1073/pnas.1218348109 PMID:23175790

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

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., 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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

Veith, G. D., & Mekenyan, O. G. (1993). A QSAR approach for estimating the aquatic toxicity of soft electrophiles. *Quantitative Structure-Activity Relationships*, *12*(4), 349–356. doi:10.1002/qsar.19930120402

Vela, A., & Gázquez, J. L. (1990). A relationship between the static dipole polarizability, the global softness, and the Fukui function. *Journal of the American Chemical Society*, *112*(4), 1490–1492. doi:10.1021/ja00160a029

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

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

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

Vermeire, T. G., Jager, D. T., Bussian, B., Devillers, J., den Haan, K., Hansen, B., & van der Zandt, P. T. J. et al. (1997). European Union system for the evaluation of substances (EUSES). Principles and structure. *Chemosphere*, *34*(8), 1823–1836. doi:10.1016/S0045-6535(97)00017-9 PMID:9114488

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

Veselinovic, A. M., Milosavljevic, J. B., Toropov, A., & Nikolic, G. M. (2013a). SMILES-based QSAR model for arylpiperazines as high-affinity 5-HT1A receptor ligands using CORAL. *European Journal of Pharmaceutical Sciences*, *48*(3), 532–541. doi:10.1016/j.ejps.2012.12.021 PMID:23287365

Veselinovic, A. M., Milosavljevic, J. B., Toropov, A., & Nikolic, G. M. (2013b). SMILES-Based QSAR Models for the Calcium Channel-Antagonistic Effect of 1,4-Dihydropyridines. *Archiv der Pharmazi Chemistry in Life Sciences*, *346*(2), 134–139. doi:10.1002/ardp.201200373 PMID:23280520

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

Vijayaraj, R., Subramanian, V., & Chattaraj, P. K. (2009). Comparison of global reactivity descriptors calculated using various density functionals: A QSAR perspective. *Journal of Chemical Theory and Computation*, *5*(10), 2744–2753. doi:10.1021/ct900347f

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

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 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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.

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

Vinogradov, S. V., Bronich, T. K., & Kabanov, A. V. (2002). Nanosized cationic hydrogels for drug delivery: Preparation, properties and interactions with cells. *Advanced Drug Delivery Reviews*, *54*(1), 135–147. doi:10.1016/S0169-409X(01)00245-9 PMID:11755709

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

von der Krammer, F., Ferguson, P. L., Holden, P. A., Masion, A., Rogers, K. R., & Klaine, S. J. et al.. (2012). Analysis of engineered nanomaterials in complex matrices (environment and biota): General consideractions and conceptual case studies. *Environmental Toxicology and Chemistry*, *31*(1), 32–49. doi:10.1002/etc.723 PMID:22021021

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

Voskresensky, O. N., & Levitsky, A. P. (2002). QSAR aspects of flavonoids as a plentiful source of new drugs. *Current Medicinal Chemistry*, *9*(14), 1367–1383. doi:10.2174/0929867023369790 PMID:12132993

Vracko, M. (1997). A study of structure-carcinogenic potency relationship with artificial neural networks. The using of descriptors related to geometrical and electronic structures. *Journal of Chemical Information and Computer Sciences*, *37*(6), 1037–1043. doi:10.1021/ci970231y

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

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

Walczak, B., Daszykowski, M., & Stanimirova, I. (2009). Robust methods in QSAR. In T. Puzyn, J. Leszczynski, & M. T. D. Cronin (Eds.), *Recent advances in QSAR studies. Challenges and advances in computational chemistry and physics*. Dordrecht, The Netherlands: Springer.

Walker, J. D., Jaworska, J., Comber, J. 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

Walker, N. J., & Bucher, J. R. (2009). A 21st century paradigm for evaluating the health hazards of nanoscale materials? *Toxicological Sciences*, *110*(2), 251–254. doi:10.1093/toxsci/kfp106 PMID:19468057

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

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

Wang, J., Byrne, J. D., Napier, M. E., & DeSimone, J. M. (2011). More effective nanomedicines through particle design. *Small*, *7*(14), 1919–1931. doi:10.1002/smll.201100442 PMID:21695781

Wang, K., Pei, H., Huang, B., Zhu, X., Zhang, J., & Zhou, B. et al. (2013). The expression of ABC efflux pump, Rv1217c–Rv1218c, and its association with multidrug resistance of *Mycobacterium tuberculosis* in China. *Current Microbiology*, *66*(3), 222–226. doi:10.1007/s00284-012-0215-3 PMID:23143285

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, 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

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, 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

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

Wania, F. (2006). Potential of degradable organic chemicals for absolute and relative enrichment in the *Arctical*. *Environmental Science & Technology*, *40*(2), 569–577. doi:10.1021/es051406k PMID:16468404

Wania, F., & Daly, G. L. (2002). Estimating the contribution of degradation in air and deposition to the deep sea to the global loss of PCBs. *Atmospheric Environment*, *36*(36-37), 5581–5593. doi:10.1016/S1352-2310(02)00693-3

Wania, F., & Mackay, D. (1995). A global distribution model for persistent organic-chemicals. *The Science of the Total Environment*, *160-61*, 211–232. doi:10.1016/0048-9697(95)04358-8

Wan, J., Zhang, L., Yang, G., & Zhan, C.-G. (2004). Quantitative structure-activity relationship for cyclic imide derivatives of protoporphyrinogen oxidase inhibitors: A study of quantum chemical descriptors from density functional theory. *Journal of Chemical Information and Computer Sciences*, *44*(6), 2099–2105. doi:10.1021/ci049793p PMID:15554680

Warheit, D. B., Sayes, C. M., Reed, K. L., & Swain, K. A. (2008). Health effects related to nanoparticle exposures: Environmental, health and safety considerations for assessing hazards and risks. *Pharmacology & Therapeutics*, *120*(1), 35–42. doi:10.1016/j.pharmthera.2008.07.001 PMID:18703086

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

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

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

Wasaznik, A., Grinholc, M., & Bielawski, K. P. (2008). Active efflux as the multidrug resistance mechanism. *Postepy higieny i medycyny doswiadczalnej (Online)*, *63*, 123–133. PMID:19373193

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

Weaver, S., & Paul Gleeson, M. (2008). The importance of the domain of applicability in QSAR modeling. *Journal of Molecular Graphics & Modelling*, *26*(8), 1315–1326. doi:10.1016/j.jmgm.2008.01.002 PMID:18328754

Wegmann, F., Cavin, L., MacLeod, M., Scheringer, M., & Hungerbuhler, K. (2009). The OECD software tool for screening chemicals for persistence and long-range atmospheric transport potential. *Environmental Modelling & Software*, *24*(2), 228–237. doi:10.1016/j.envsoft.2008.06.014

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

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

Weininger, D. (1990) SMILES. 3. Depict. Graphical depiction of chemical structures. *Journal of Chemical Information and Computer Sciences, 30*(3), 237-243.

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

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

Weissleder, R., Kelly, K., Sun, E. Y., Shtatland, T., & Josephson, L. (2005). Cell-specific targeting of nanoparticles by multivalent attachment of small molecules. *Nature Biotechnology*, *23*(11), 1418–1423. doi:10.1038/nbt1159 PMID:16244656

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

Westerhoff, P., & Nowack, B. (2013). Searching for global descriptors of engineered nanomaterial fate and transport in the environment. *Accounts of Chemical Research*, *46*(3), 844–853. doi:10.1021/ar300030n PMID:22950943

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

White, R. M., Sessa, A., Burke, C., Bowman, T., Leblanc, J., & Ceol, C. et al. (2008). Transparent adult zebrafish as a tool for in vivo transplantation analysis. *Cell Stem Cell*, *2*(2), 183–189. doi:10.1016/j.stem.2007.11.002 PMID:18371439

Whitley, D. (1994). A genetic algorithm tutorial. *Statistics and Computing*, *4*(2), 65–85. doi:10.1007/BF00175354

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

Wieland, T. (1997). Combinatorics of combinatorial chemistry. *Journal of Mathematical Chemistry*, *21*(2), 141–157. doi:10.1023/A:1019166201637

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

Wiener, H. (1948). Vapor pressure-temperature relationships among the branched paraffin hydrocarbons. *The Journal of Physical and Colloid Chemistry*, *52*(2), 425–430. doi:10.1021/j150458a014 PMID:18906414

Wigner, E. P. (1960). The unreasonable effectiveness of mathematics in the natural sciences. Richard courant lecture in mathematical sciences delivered at New York University, May 11, 1959. *Communications on Pure and Applied Mathematics*, *13*(1), 1–14. doi:10.1002/cpa.3160130102

Wijnhoven, S. W. P., Peijnenburg, W. J. G. M., Herberts, C. A., Hagens, W. I., Oomen, A. G., & Heugens, E. H. W. et al. (2009). Nano-silver – A review of available data and knowledge gaps in human and environmental risk assessment. *Nanotoxicology*, *3*(2), 109–138. doi:10.1080/17435390902725914

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

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

Wishart, D. S., Knox, C., Guo, A. C., Cheng, D., Shrivastava, S., & Tzur, D. et al. (2008). DrugBank: A knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Research*, *36*(Database), 901–906. doi:10.1093/nar/gkm958 PMID:18048412

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., & Eriksson, L. (1995). Chemometric methods. In H. van de Waterbeemd (Ed.), Molecular design (pp. 312–317). Weinheim, Germany: VCH.

Wold, S., & Eriksson, L. (1995). Statistical validation of QSAR results. In Chemometrics methods in molecular design (pp. 309-318). Weinheim, Germany: Wiley-VCH.

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.

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

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., & 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

Wold, S., Sjostrom, M., & Eriksson, L. (2001). PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, *58*(2), 109–130. doi:10.1016/S0169-7439(01)00155-1

Wolff, M. E. (1955). Therapeutic agents. In *Burger's medicinal chemistry and drug discovery* (Vol. 4). New York: John Wiley & Sons.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

Wood, A. J., Gold, H. S., & Moellering, R. C. Jr. (1996). Antimicrobial-drug resistance. *The New England Journal of Medicine*, *335*(19), 1445–1453. doi:10.1056/NEJM199611073351907 PMID:8875923

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

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

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

Worachartcheewan, A., Nantasenamat, C., Isarankura-Na-Ayudhya, C., & Prachayasittikul, V. (2014). QSAR study of H1N1 neuraminidase inhibitors from influenza a virus. *Letters in Drug Design & Discovery*, *11*(4), 420–427. doi:10.2174/15701808113106660085

Worth, A. P., Bassan, A., Gallegos, A., Netzeva, T. I., Patlewicz, G., & Pavan, M. et al. (2005). *The characterisation of (quantitative) structure-activity relationships: preliminary guidance.* Ispra, Italy: European Commission, Joint Research Centre.

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

Xiang, Z. (2006). Advances in homology protein structure modeling. *Current Protein & Peptide Science*, *7*(3), 217–227. doi:10.2174/138920306777452312 PMID:16787261

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

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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

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

Yan, A., & Gasteiger, J. (2003). Prediction of aqueous solubility of organic compounds based on a 3D structure representation. *Journal of Chemical Information and Computer Sciences*, *43*(2), 429–434. doi:10.1021/ci025590u PMID:12653505

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

Yan, D., Jiang, X., Xu, S., Wang, L., Bian, Y., & Yu, G. (2008). Quantitative structure-toxicity relationship study of lethal concentration to tadpole (Bufo vulgaris formosus) for organophosphorous pesticides. *Chemosphere*, *71*(10), 1809–1815. doi:10.1016/j.chemosphere.2008.02.033 PMID:18395243

Yang, G.-F., & Huang, X. (2006). Development of quantitative structure-activity relationships and its application in rational drug design. *Current Pharmaceutical Design*, *12*(35), 4601–4611. doi:10.2174/138161206779010431 PMID:17168765

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, W., & Mortier, W. J. (1986). The use of global and local molecular parameters for the analysis of the gas-phase basicity of amines. *Journal of the American Chemical Society*, *108*(19), 5708–5711. doi:10.1021/ja00279a008 PMID:22175316

Yang, W., & Parr, R. G. (1985). Hardness, softness, and the fukui function in the electronic theory of metals and catalysis. *Proceedings of the National Academy of Sciences of the United States of America*, *82*(20), 6723–6726. doi:10.1073/pnas.82.20.6723 PMID:3863123

Yang, W., Parr, R. G., & Pucci, R. (1984). Electron density, Kohn– Sham frontier orbitals, and Fukui functions. *The Journal of Chemical Physics*, *81*(6), 2862–2863. doi:10.1063/1.447964

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

Yap, C. W. (2011). PaDEL-Descriptor: An open source software to calculate molecular descriptors and fingerprints. *Journal of Computational Chemistry*, *32*(7), 1466–1474. doi:10.1002/jcc.21707 PMID:21425294

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

Yilmaz, S., Altinkanat-Gelmez, G., Bolelli, K., Guneser-Merdan, D., Over-Hasdemir, M. U., & Yildiz, I. et al. (2014). Pharmacophore generation of 2-substituted benzothiazoles as AdeABC efflux pump inhibitors in *A. baumannii*. *SAR and QSAR in Environmental Research*, *25*(7), 551–563. doi:10.1080/1062936X.2014.919357 PMID:24905472

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Yokozawa, T., Chen, C. P., Dong, E., Tanaka, T., Nonaka, G.-I., & Nishioka, I. (1998). Study on the inhibitory effect of tannins and flavonoids against the 1,1-diphenyl-2-picrylhydrazyl radical. *Biochemical Pharmacology*, *56*(2), 213–222. doi:10.1016/S0006-2952(98)00128-2 PMID:9698075

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

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

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

Yum, S., Xu, Y., Piao, S., Sim, S.-H., Kim, H.-M., & Jo, W.-S. et al. (2009). Crystal structure of the periplasmic component of a tripartite macrolide-specific efflux pump. *Journal of Molecular Biology*, *387*(5), 1286–1297. doi:10.1016/j.jmb.2009.02.048 PMID:19254725

Yuranova, T., Laub, D., & Kiwi, J. (2007). Synthesis activity and characterization of textiles showing self-cleaning activity under daylight irradiation. *Catalysis Today*, *122*(1-2), 109–117. doi:10.1016/j.cattod.2007.01.040

Zahr, A. S., de Villiers, M., & Pishko, M. V. (2005). Encapsulation of drug nanoparticles in self-assembled macromolecular nanoshells. *Langmuir*, *21*(1), 403–410. doi:10.1021/la0478595 PMID:15620331

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

Zechini, B., & Versace, I. (2009). Inhibitors of multidrug resistant efflux systems in bacteria. *Recent Patents on Anti-infective Drug Discovery*, *4*(1), 37–50. doi:10.2174/157489109787236256 PMID:19149695

Zhang, B. T., Zheng, X., Li, H. F., & Lin, J. M. (2013). Application of carbon-based nanomaterials in sample preparation: A review. *Analytica Chimica Acta*, *784*, 1–17. doi:10.1016/j.aca.2013.03.054 PMID:23746402

Zhang, H., Ji, Z., Xia, T., Meng, H., Low-Kam, C., & Liu, R. et al. (2012). Use of metal oxide nanoparticle band gap to develop a predictive paradigm for oxidative stress and acute pulmonary inflammation. *ACS Nano*, *6*(5), 4349–4368. doi:10.1021/nn3010087 PMID:22502734

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, L., Hao, G.-F., Tan, Y., Xi, Z., Huang, M.-Z., & Yang, G.-F. (2009). Bioactive conformation analysis of cyclic imides as protoporphyrinogen oxidase inhibitor by combining DFT calculations, QSAR and molecular dynamic simulations. *Bioorganic & Medicinal Chemistry*, *17*(14), 4935–4942. doi:10.1016/j.bmc.2009.06.003 PMID:19540767

Zhang, L., & Ma, S. (2010). Efflux pump inhibitors: A strategy to combat p-glycoprotein and the NorA multidrug resistance pump. *ChemMedChem*, *5*(6), 811–822. doi:10.1002/cmdc.201000006 PMID:20373322

Zhang, L., Zhou, P., Yang, F., & Wang, Z. (2007). Computer-based QSARs for predicting mixture toxicity of benzene and its derivatives. *Chemosphere*, *67*(2), 396–401. doi:10.1016/j.chemosphere.2006.09.018 PMID:17184822

Zhang, S. G., Lei, W., Xia, M. Z., & Wang, F. Y. (2005). QSAR study on N-containing corrosion inhibitors: Quantum chemical approach assisted by topological index. *Journal of Molecular Structure THEOCHEM*, *732*(1-3), 173–182. doi:10.1016/j.theochem.2005.02.091

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

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, 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

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

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

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

Zheng, F., Zhan, M., Huang, X., Hammed, M. D. M. A., & Zhan, C.-G. (2014). Modeling *in vitro* inhibition of butyrylcholinesterase using molecular docking, multilinear regression and artificial neural network approaches. *Bioorganic & Medicinal Chemistry*, *22*(1), 538–549. doi:10.1016/j.bmc.2013.10.053 PMID:24290065

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

Zhou, H., Mu, Q., Gao, N., Liu, A., Xing, Y., & Gao, S. et al. (2008). A nano-combinatorial library strategy for the discovery of nanotubes with reduced protein-binding, cytotoxicity, and immune response. *Nano Letters*, *8*(3), 859–865. doi:10.1021/nl0730155 PMID:18288815

Zhou, K., Wang, R., Xu, B., & Li, Y. (2006). Synthesis, characterization and catalytic properties of CuO nanocrystals with various shapes. *Nanotechnology*, *17*(15), 3939–3943. doi:10.1088/0957-4484/17/15/055

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

Zhu, X., Zhu, L., Duan, Z., Qi, R., Li, Y., & Lang, Y. (2008). Comparative toxicity of several metal oxide nanoparticle aqueous suspensions to Zebrafish (Danio rerio) early developmental stage. *Journal of Environmental Science and Health, Part A, 43*(3), 278-284.

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

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

Zhu, X., Chang, Y., & Chen, Y. (2010). Toxicity and bioaccumulation of TiO_2 nanoparticle aggregates in Daphnia magna. *Chemosphere*, *78*(3), 209–215. doi:10.1016/j.chemosphere.2009.11.013 PMID:19963236

Zhu, X., Tian, S., & Cai, Z. (2012). Toxicity assessment of iron oxide nanoparticles in zebrafish (Danio rerio) early life stages. *PLoS ONE*, *7*(9), e46286. doi:10.1371/journal.pone.0046286 PMID:23029464

Zhu, X., Zhu, L., Chen, Y., & Tian, S. (2009). Acute toxicities of six manufactured nanomaterial suspensions to *Daphnia magna. Journal of Nanoparticle Research*, *11*(1), 67–75. doi:10.1007/s11051-008-9426-8

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Zingales, B., Andrade, S. G., Briones, M. R. S., Campbell, D. A., Chiari, E., & Fernandes, O. et al. (2009). A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: Second revision meeting recommends TcI to TcVI. *Memorias do Instituto Oswaldo Cruz*, *104*(7), 1051–1054. doi:10.1590/S0074-02762009000700021 PMID:20027478

Zukowska, B., Breivik, K., & Wania, F. (2006). Evaluating the environmental fate of pharmaceuticals using a level III model based on poly-parameter linear free energy relationships. *Environmental Science & Technology*, *359*, 177–187. PMID:16009398

Zupan, J. (1994). Introduction to artificial neural network (ANN) methods: What they are and how to use them. *Acta Chimica Slovenica*, *41*, 327–327.

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

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

About the Contributors

Kunal Roy is an Associate Professor in the Department of Pharmaceutical Technology, Jadavpur University, Kolkata, India, and a visiting Fellow in the Manchester Institute of Biotechnology, University of Manchester, United Kingdom. He is an Associate Editor of the Springer journal, *Molecular Diversity*, and a member of the Editorial Advisory Board of *European Journal of Medicinal Chemistry* (Elsevier), *Chemical Biology and Drug Design* (Wiley) and *Expert Opinion on Drug Discovery* (Informa). The field of his research interest is QSAR and Molecular Modeling. Dr. Roy has published more than 225 research papers in refereed journals.

* * *

Nilanjan Adhikari, a senior research fellow of Council of Scientific and Industrial Research (CSIR), New Delhi, is pursuing research in Dept. of Pharmaceutical Technology, Jadavpur University, under the guidance of Prof. (Dr.) Tarun Jha. His research area includes designing and synthesis of anticancer small molecules. He has published 16 research articles in different reputed journals.

Antreas Afantitis, PhD, MBA, is the director of NovaMechanics Ltd. Dr. Antreas Afantitis is an experienced consultant in chemoinformatics and applied drug discovery projects. Dr. Afantitis has a strong scientific background in the field of chemoinformatics, bioinformatics, and medicinal chemistry. He has been included in the most productive authors in the field of Chemoinformatics (bibliometric analysis performed by Prof. Peter Willet) and according to Google Scholar (2014) his h-index is 20. He is a reviewer in leading scientific journals, member of Editorial Boards (more than 4), and an evaluator of international competitive grants. Recently, he has joined *Combinatorial Chemistry and High Throughput Screening* as the Section Editor for Combinatorial/Medicinal Chemistry. He has developed and applied several virtual screening methodologies to drug discovery problems and based on the results three patents were filled and several articles (over 40) have been published in top scientific journals. Dr. Afantitis is actively involved (PI or co-PI) in several FP7 EU projects, Cost Actions, Eureka, and Nationally Funded R&D projects.

Daniel E. Bacelo graduated in Chemistry at Universidad de Buenos Aires in 1994 and received his PhD in Physical Chemistry at University of Puerto Rico-Rio Piedras campus in 1998 under the direction of Y. Ishikawa in the area of Theoretical Chemistry. He was Full Professor of Chemistry at Universidad Metropolitana of Puerto Rico and researcher at Universidad Nacional de la Patagonia San Juan Bosco, Argentina. Currently, he is Professor of Chemistry at Universidad de Belgrano and researcher at National

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Scientific Research Council (CONICET, Argentina). His research interests are, among others, studies of bioactive molecules using accurate quantum chemistry calculations, atomistic modeling of metallic nanostructures, and molecular activity determination using QSAR.

Emilio Benfenati is head of the Laboratory of Environmental Chemistry and Toxicology at the Mario Negri Institute, Milan, Italy, since 1997. In the laboratory about 35 researchers are active. Previously, he has been researcher at Stanford University, California, USA (in that period he also had a collaborative research at the Berkeley University, California, USA). He coordinated 16 European projects (including CAESAR, CALEIDOS, ANTARES, PROSIL, ToxBank, IMAGETOX) and participated to 19 others. Many of them are on toxicity and environmental modelling. His research activities include toxicity and environmental modelling, molecular descriptors, QSARs, toxicity prediction, environmental management, characterisation and assessment of contaminants, risk assessment; development of QSAR models; analysis of environmental and food samples for pollutants such as dioxins, PCB, PAH, pesticides, endocrine disruptors, industrial pollutants; environmental assessment. He is author or co-author of about 250 papers in international journals and edited a few books.

Angelo Carotti graduated in Chemistry at the University of Camerino and became Full Professor of Medicinal Chemistry in 1990 at the Faculty of Pharmacy of the University of Bari. He co-authored 3 patents and more than 210 publications concerning the molecular design, synthesis, and biological evaluation of bioactive agents against neurodegenerative and cancer diseases. Prof. Carotti was the director of the European School of Medicinal Chemistry of Urbino (1999-2004) and of the Doctoral School of "Pharmaceutical Sciences" (2007-2012). Currently, he coordinates the Doctoral course of Biomolecular Pharmaceutical and Medical Sciences. In 2013, he was awarded the "Musajo" medal for his scientific and professional achievements by the Division of Medicinal Chemistry of the Italian Chemical Society.

Eduardo A. Castro performed his PhD work at the Physics Department of La Plata National University from 1971-72, working under supervision of Manuel Sorarrain. After that, he became a research scientist at the Theoretical and Applied Research Institute located at La Plata National University, where he founded the Group for Theoretical Chemistry in 1974. At the present time, he is as a Superior Researcher from the National Scientific and Technical Research Council of Argentina. He is reviewer of several journals, member of the editorial board of a large number of scientific journals, and editor and co-editor of preeminent journals and publishers from all continents. Professor Castro´s scientific contributions encompass a rather large number of fields such as electronic molecular structure, physical organic chemistry, QSAR-QSPR theory, science education, administration of scientific organizations, science communication, supramolecular chemistry, solid structure theory, just to mention a few.

Pratim Kumar Chattaraj obtained his BSc and MSc degrees in Chemistry from Burdwan University and his PhD degree from Indian Institute of Technology Bombay. Presently, he is a professor in the Department of Chemistry, Indian Institute of Technology Kharagpur, and also the Convener of the Center for Theoretical Studies there. He was a postdoctoral research associate in the University of North Carolina (Chapel Hill). His research interests include density functional theory, the theory of chemical reactivity, aromaticity in metal clusters, ab initio calculations, quantum trajectories, and nonlinear dynamics. Professor Chattaraj is a member of the editorial board of J. *Mol. Struct. (Theochem), J. Chem. Sci.*, and *Ind. J. Chem. A,* among others. He was the Head of the Department and a

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

council member of the Chemical Research Society of India. He is a Fellow of the World Academy of Science, the Indian Academy of Sciences (Bangalore), the Indian National Science Academy (New Delhi), the National Academy of Sciences, India (Allahabad), and the West Bengal Academy of Science and Technology. He is a J. C. Bose National Fellow. He has published around 300 scientific publications in international journals.

M. Natália D. S. Cordeiro obtained her first degree in Chemistry at the Faculty of Sciences, University of Porto, Portugal, followed by a PhD in Monte Carlo Simulations of Ionic Aqueous Solutions. In 1996, she became a Professor in Theoretical Chemistry at the University of Porto, and she is now the leader of a group working in chemoinformatics and materials modelling. Her research interests have been focused on the properties of ionic solutions, fluid interfaces, polymers, nanomaterials, heterogeneous catalysis, reaction dynamics and mechanisms, and lately QSAR studies applied to drug discovery or toxicology problems.

Rudra Narayan Das is currently carrying out research at the Department of Pharmaceutical Technology, Jadavpur University (India) on the exploration of diagnostic features of ETA indices towards modeling various activity, property and toxicology endpoints. He completed his B. Pharmacy (2009) and M. Pharmacy (2011) degrees from the same university. His M. Pharm. project also involved studies on chemoinformatics. He has experience in developing predictive classification and regression-based chemometric models on different endpoints data viz. ecotoxicity involving indicator organisms (Tetrahymena, Daphnids, Pimephales, Vibrio, etc.), rodent toxicity, aqueous solubility, partition coefficient, molar refractivity, anti-malarial activity, etc. employing varying group of chemicals involving simple organic molecules, complex drugs, and novel ionic liquids. He has published about 15 research and review articles in peer-reviewed journals and one book chapter as well.

Siavoush Dastmalchi graduated as Doctor of Pharmacy from Tabriz University of Medical Sciences (TUMS). Then he moved to Sydney where he received his PhD from the Faculty of Pharmacy at the University of Sydney in 2002. Since then he has worked as a full academic in the Medicinal Chemistry Department at the School of Pharmacy, TUMS, teaching medicinal chemistry, instrumental drug analysis and bioinformatics to graduate and postgraduate students. He is currently the Director of the Biotechnology Research Centre at TUMS where he leads his research team mainly with interests in molecular modelling, structural biology, and chemo-bioinformatics for their application to drug discovery.

Omar Deeb received his MSc (Analytical Chemistry) in 1986 and his PhD (Molecular Modelling and Drug Design) in 1993 from the Hebrew University of Jerusalem. This was followed by post-doctoral research at Chemometrics Laboratory, Orleans University, France, on QSAR of different pharmaceutical compounds. He joined Al-Quds University in 1995 and is currently a Professor in the faculty of Pharmacy. He is the group leader of computational chemistry/drug design laboratory. His research interests include QSAR and molecular modelling of pharmaceutical compounds, hydrogen bonding interactions, and proton transfer. His research achievements include QSAR of different groups of pharmaceutical compounds which leads to designing new drugs, application of new techniques such as flip regression for handling the symmetry of different systems, and the use of the orbital nodal angle descriptors in QSAR. Most of his publications including key articles, chapters in books, proceedings and oral presentations at international meetings related to QSAR.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Trong-Nhat Do was born in District 1, Ho Chi Minh City, Vietnam. Do earned his BSPharm degree at the University of Medicine and Pharmacy at Ho Chi Minh City, Vietnam in 2014. He worked as a staff member of the non-governmental organization VietHope in 2008-2011 and as a board member of Vietnam Student Development Fund (VSDF) in 2011-2012. He served in the Management Board of *Pharmagazine* – the official bilingual magazine of the Youth Union at the School of Pharmacy in 2012-2013, taking charge of the PharmInfo column. He has presented his research at some scientific conference meetings.

Quoc-Hiep Dong earned his Bachelor of Science degree in Pharmacy from the University of Medicine and Pharmacy, Ho Chi Minh City, Vietnam, in 2014. His thesis was titled "Modeling and Virtual Screening of P1–P3 Macrocyclic Analogs as NS3/4A Serine Protease Inhibitors". Currently, he is an assistant researcher at Department of Medicinal Chemistry, University of Medicine and Pharmacy at Ho Chi Minh City, Vietnam.

Mariene Helena Duarte was born in Santo Antonio do Amparo (Brazil) and studied undergraduate Chemistry and did her MSc in Agrochemistry (Biochemistry) at the Federal University of Lavras (UFLA). She is currently developing her PhD at UFLA under supervision of Prof. Matheus P. Freitas, where she has been working with the QSAR modeling of bioactivities of compounds against neglected diseases. Department of Chemistry, Federal University of Lavras.

Pablo R. Duchowicz studied Physical-Chemistry from 1996-2003 at the Faculty of Exact Sciences, Chemistry Department of La Plata National University (UNLP), Buenos Aires, Argentina. His PhD work on "Physicochemical and Biological Applications of the QSPR-QSAR Theory," was performed at the Research Institute of Theoretical and Applied Physical-Chemistry (INIFTA) located at La Plata, under the supervision of Profs. Eduardo A. Castro and Francisco M. Fernández, from 2003-2005. In 2006, he obtained a post-doctoral scholarship to work on "ab initio Direct Kinetics and Molecular Dynamics Studies for Halogenated Germanes and Related Species" at INIFTA, under the supervision of Dr. Carlos J. Cobos and Prof. Adela Croce. Since 2007, he is a member of the Scientific Researcher Career of the National Research Council of Argentina, performing research work at INIFTA.

Matheus Puggina de Freitas was born in Itapira (Brazil) and studied Chemistry (BSc and PhD) at the State University of Campinas (UNICAMP), following a postdoctoral stay at the same university. He worked as researcher in two pharmaceutical companies, and in 2005, he started as Adjunct Professor at the Federal University of Lavras (UFLA). He is currently associate Professor of Organic Chemistry at UFLA, where he has research interests in conformational analysis of small molecules, NMR spectroscopy, computational chemistry, and medicinal chemistry.

Mohammad Goodarzi is currently a post-doctoral researcher at BIOSYST-MeBioS, Faculty of Bioscience Engineering, K.U. Leuven, Belgium. He obtained his PhD in 2012 on feature selection and modeling chemometrics techniques in QSAR/QSPR studies at the Department of Analytical Chemistry and Pharmaceutical Technology, Vrije Universiteit Brussel, Belgium. His main focus is devoted to the development and application of Chemometrics in Spectroscopy, Hyperspectral Imaging, and Omics Data Analysis of plant and animal origin. He is co-author of over 90 scientific publications.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Ashutosh Gupta received his MSc from IIT Bombay in 1998 and PhD from Purvanchal University, Jaunpur, India, in 2007. In 2003, he was hired as Assistant Professor by Udai Pratap (Autonomous) College, Varanasi, UP, India, and is currently working as Senior Assistant Professor in the same college. He has worked extensively as Visiting Fellow to different institutes within and outside India. His area of interest is in computational chemistry with an emphasis on reaction mechanism determination and application of conceptual density functional theory in the elucidation of different reaction pathways. He has also contributed in the area of QSAR and their application in the determination of different properties related to eco-toxicity.

Shikhar Gupta has done a Master's in Organic Chemistry and worked as Research Intern in "C.I.MAP" (Central Institute of Medicinal & Aromatic Plants), Lucknow. Afterwards, he joined as JRF in the Indo-Finland project "Structural Bioinformatics for Effective and Efficient Drug Discovery at NIPER" (National Institute of Pharmaceutical Education & Research) and finished his PhD in Computational Chemistry/Molecular Modeling in 2011.

Amit Kumar Halder, a senior research fellow of Council of Scientific and Industrial Research (CSIR), New Delhi, is pursuing research in Dept. of Pharmaceutical Technology, Jadavpur University, under the guidance of Prof. (Dr.) Tarun Jha. His research area includes designing and synthesis of anticancer small molecules. He has published 19 research articles in different reputed journals.

Maryam Hamzeh-Mivehroud is an assistant professor who graduated as Doctor of Pharmacy from Tabriz University of Medical Sciences in 2004 and received her PhD in Medicinal Chemistry from this university in 2011. Since then, she has worked as a full academic member in the Medicinal Chemistry Department at the School of Pharmacy and teaches medicinal chemistry at the undergraduate and postgraduate levels. Her main research interests are focused on QSAR and molecular modelling based on various description selection algorithms and machine learning methods. She is also professional in working with modelling and statistical softwares such as GROMACS, GOLD, AMBER, GOLPE, MOE, etc.

Karolina Jagiello, PhD, Eng, got the Diploma in 2005 in Biotechnology and PhD in 2010 in Chemistry (Gdansk University of Technology, Poland). After graduating, she started collaboration with Laboratory of Environmental Chemometrics leading by Prof. Tomasz Puzyn firstly as a stipendist of Foundation of Polish Science in the frame of grant coordinated by Prof. Puzyn. Since 2013, Dr. Jagiello works as Adjunct at the University of Gdansk. Dr. Jagiello is young scientist with expertise in application of computational techniques in environmental risk assessment. She is an author or co-author of six articles and four chapters including those in leading journals (e.g. *Environmental Science and Technology*).

Tarun Jha, a faculty member of Dept. of Pharmaceutical Technology, Jadavpur University, has supervised eight PhD students and guided five research projects funded by different organizations. He has published more than 70 research articles in different reputed journals. His research area includes designing and synthesis of anticancer small molecules. He is one of the E.C. members of National Board of Accreditation (NBA), New Delhi.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Supratik Kar is a researcher in the Department of Pharmaceutical Technology, Jadavpur University, Kolkata, India. He has completed his BPharm (2008) and MPharm (2010) degrees from Jadavpur University, securing first position in both degrees. The field of his research interest is QSAR and Molecular Modeling. He is actively working in the development of in silico models for assessing the human and environmental risk of novel chemicals and pharmaceuticals. He has an experience of about six years in the area of QSAR and published 28 research articles in peer-reviewed journals (https://sites.google.com/site/visitsupratikkar/home/publications). He has visited Laboratory of Environmental Cheminformatics of University of Gdansk, Poland, under NanoBRIDGES project funded by Marie Curie research fellowship, European Union (http://nanobridges.eu/supratik-kar-ju-in-university-of-gdansk-ug-in-gdansk/).

Lemont B. Kier is an Emeritus Professor of Medicinal Chemistry at the Virginia Commonwealth University, Richmond, Virginia, USA. Dr. Kier received a BS in Pharmacy from The Ohio State University, USA, and a PhD in Medicinal Chemistry from the University of Minnesota, USA. Dr. Kier was an early developer of drug design strategies using quantitative structure-activity information. This included molecular orbital theory to predict preferred conformations and numerical descriptors encoding molecular structure and electronic properties. More recent research has developed new approaches to the modeling of molecular structure using cellular automata. These include studies of dynamic processes leading to theories of anesthetic action, sleep, aging, and nerve conduction. Close to 300 papers and 8 books have revealed these contributions. Dr. Kier is active in medicinal chemistry, life sciences, and nurse anesthesia courses at the Virginia Commonwealth University.

Valeria V. Kleandrova graduated in 2010 in Engineering in Technology of Fermentative Production and Wine Industry, Moscow State University of Food Production, Russia. Since 2011, she has been working as young researcher under the supervision of Prof. Dr. Maria Natália Dias Soeiro Cordeiro, at REQUIMTE/Department of Chemistry and Biochemistry, University of Porto, Portugal. Her principal research interests are devoted to the application of chemoinformatics methods in areas such as Medicinal Chemistry, Nutrition, Nanotoxicology, Nanomedicine, and Environmental Sciences. She has been co-author of 20 publications in journals of international impact.

Duy-Phong Le graduated Bachelor of Pharmacy Diploma (BS in Pharmacy) from Faculty of Pharmacy, University of Medicine and Pharmacy at Ho Chi Minh City, Vietnam, in 2013. His research interest is computer-aided drug discovery and development. Currently, he is an assistant researcher at Department of Medicinal Chemistry, University of Medicine and Pharmacy at Ho Chi Minh City, Vietnam.

Minh-Tri Le graduated Bachelor of Pharmacy Diploma (BS in Pharmacy) from Faculty of Pharmacy, University of Medicine and Pharmacy at Ho Chi Minh City, Vietnam, in 1984, and earned his PhD degree in Pharmaceutical Sciences in 1997. Currently, Assoc. Prof. Le serves as Head of Department of Medicinal Chemistry at Faculty of Pharmacy, University of Medicine and Pharmacy at Ho Chi Minh City, Vietnam.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Danuta Leszczynska is a Professor of Environmental Engineering and Environmental Science in the Department of Civil and Environmental Engineering at the Jackson State University. She completed her education at the Technical University of Wroclaw, Poland, with degrees in Chemical Engineering (BS), Polymer Science and Technology (MS), and Environmental Engineering (PhD). Her current research concentrates on the properties, impact, and incidental interactions of engineered nanomaterials in environment. She is author and co-author of over 200 peer-reviewed publications. She is Fulbright Scholar (2005), member of National Academy of Inventors (since 2011), and she was recognized by her Alma Mater as a Distinguish Alumni (2012).

Jerzy Leszczynski, Professor of Chemistry and President's Distinguished Fellow at the Jackson State University, directs Interdisciplinary Nanotoxicity NSF CREST Center. Dr. Leszczynski is a computational quantum chemist whose vast areas of interest include nature of chemical bonds, theoretical predictions of molecular potential energy surfaces and vibrational spectra, structures and properties of molecules with heavy elements, properties and structure of DNA fragments, and characteristics of nanomaterials. He also applies computational chemistry methods to environmental problems, surface chemistry and atmospheric chemistry. He has published over 800 referred papers and 60 book chapters. Dr. Leszczynski is editor of a book series: *Computational Chemistry: Reviews of Current Trends* (World Scientific); editor of a book series *Challenges and Advances in Computational Chemistry and Physics* (Springer); editor of *Handbook of Computational Chemistry* (Springer); editor of book series *Practical Aspects of Computational Chemistry* (Springer); and series editor for *Lecture Notes in Chemistry* (Springer); and editor and member of editorial boards of eight journals.

Feng Luan is leader of the Theoretical and Computational research group at the Department of Applied Chemistry, Yantai University. The group's research focuses on computer-aided drug design and QSAR or QSPR studies of different chemical compounds, including nanomaterials. The group is focused on the development of machine learning algorithms such as support vector machines, artificial neutral networks, etc. for rational drug discovery and toxicology problems.

Thomas Mavromoustakos was born in Cyprus (1960) and completed his undergraduate studies in the Chemistry Department of National and Kapodistrian University of Athens (NKUA) (1985). He pursued graduate and post-doctoral studies in the University of Connecticut (1991). He worked as a director of Laboratory of Molecular Analysis in the National Hellenic Research Foundation (1991). He was elected as an Associate Professor of the Laboratory of Organic Chemistry in Chemistry Department of NKUA (2007) and was promoted to Full Professor (2012). For the last two years, he served as a Director of the Section II of the Chemistry Department. His research interests focus on the interactions of drugs with their site of action (membranes, DNA, receptors) and various other aspects of rational drug design (conformational analysis, docking, molecular dynamics, searching for pharmacophores, 3D QSAR, drug delivery). To utilize these aims, he is using a combination of biophysical techniques and computational analysis.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Georgia Melagraki, PhD, MBA, is an experienced consultant in chemoinformatics, applied drug discovery, and materials projects, and she has a strong scientific background in the field of computational and organic chemistry. Her research is mainly focused in drug discovery, material design, and properties prediction, and she has developed several computational tools to support decisions under uncertainty using virtual screening and informatics methods that significantly reduce costs and timescales. Her research has received funding from several competitive EU and national grants. Her scientific work has been published in several original research articles (over 40) in international peer-reviewed journals and her articles have been highlighted several times as top cited and accessed articles. According to Google Scholar, her h-index is 20 and her work has been cited several times (over 1000). She is a reviewer in leading scientific journals (more than 30) and invited member of Editorial Boards in top scientific international journals.

Nikola Minovski was born in Skopje (Republic of Macedonia) where he obtained his BSc diploma at the Faculty of Pharmacy, Ss. Cyril and Methodius University in 2005. In 2007, he moved to Slovenia and joined the PhD research program Biomedicine at University of Ljubljana (Slovenia) under the supervision of Prof. Tom Solmajer. In 2011, he completed his PhD thesis on the integration of in silico (chemometrics and molecular modeling) methods for design of novel 6-fluoroquinolone antibacterials as potential *Mycobacterium tuberculosis* DNA gyrase inhibitors. Since 2007, he is actively working as a researcher at the Laboratory for Chemometrics (group of dr. Marjana Novič), National Institute of Chemistry in Ljubljana (Slovenia). His principal research interest includes ligand-protein interactions, enzyme inhibition, antibacterial agents, QSAR, cheminformatics, and development/optimization of efficient *in silico* platforms for design and identification of novel hit candidates.

C. Gopi Mohan is currently an Associate Professor at Amrita Centre for Nanoscience and Molecular Medicine specializing in the area of Structural Bioinformatics, Chemoinformatics, Nanoinformatics and its biomedical applications. Currently, his team at centre focuses on Structure-Based Drug Design, Pharmacophore Modeling, QSAR Modeling, and Nanomaterial Modeling using different in silico techniques. He worked as a faculty member of National Institute of Pharmaceutical Education & Research (NIPER), Mohali, Punjab, and also as a post-doctoral fellow in Indian Institute of Science, Bangalore; University of Bath, UK; and CNRS research fellow at University Henri Poincare, France. He had been recipient of Indo-Finland grant for computational biology related to drug development and visited University of Helsinki and University of Turku to complete successfully this collaborative bilateral program. His protein models and other related computational work was appreciated by US-based pharmaceutical company leading to future collaboration. He was also cited as international well-known expert in the field of Structural Bioinformatics and Chemoinformatics by Synergix Ltd. United Kingdom, which was founded by Dr. N.C. Cohen, a pioneer of rational drug design.

Karel Nesmerak, born in 1974, received his PhD in Analytical Chemistry at Charles University in Prague in 2003. From 1999–2002, he worked at the Institute of Microbiology, Academy of Sciences of the Czech Republic. Since 2003, he is an assistant professor at the Department of Analytical Chemistry, Charles University in Prague. His research program focuses on the electroanalytical methods in modelling of metabolism and in the QSPR/QSAR. His studies also address the development of simple, inexpensive, and effective spectrometric and separation methods for pharmaceutical analysis. He is author/co-author of 29 papers and 5 chapters in books.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Duc-Khanh-Tho Nguyen graduated Bachelor of Pharmacy Diploma (BS in Pharmacy) from Faculty of Pharmacy, University of Medicine and Pharmacy at Ho Chi Minh City, Vietnam, in 2013. She was awarded "Odon Vallet Scholarship" by Rencontres du Vietnam (2007, 2008, 2012) and "GE Foundation Scholar Leader Scholarship" by GE Viet Nam and International Institute of Education (2009-2012).

Thi-Thanh-Lan Nguyen earned her Bachelor of Science degree in Pharmacy from the University of Medicine and Pharmacy, Ho Chi Minh City, Vietnam, in 2014. While pursuing her degree, Nguyen worked as a research associate in Drug Design Laboratory, Medicinal Chemistry Department. She conducted a research named "2D-QSAR Model and Molecular Docking Study for 11beta-Hydroxysteroid Dehydrogenase 1 Inhibitors" as her graduation thesis. Her research was supervised by Dr. Thai Khac Minh and MSc. Nguyen Thi-Thu Ha. Furthermore, she has been one of the students that had the best academic record in her class and a dynamic president of English Speaking Club at school. She has attended the exchange program between Japan and Vietnam at Takasaki University of Health and Welfare in 2013. She also achieved Daewoong Scholarship that was awarded for outstanding students in 2012.

Thuy-Viet-Phuong Nguyen earned her Pharmacist degree from the University of Medicine and Pharmacy in Ho Chi Minh City, Vietnam, in 2003. She was awarded a Master's degree in Pharmaceutics and Pharmaceutical Technology in 2008. Currently, she is a third-year PhD student in the School of Chemistry, University of Wollongong, Australia. Her research interested in is computer-aided drug discovery and development. Her thesis focuses on antiviral drug design, with particular emphasis on the Chikungunya virus, using computational approaches such as molecular docking and molecular dynamics simulations in developing the inhibitors for the virus.

Orazio Nicolotti received the degree (110/110 cum laude) in Chemistry and Pharmaceutical Technology in 1997 and the PhD in Pharmaceutical Science in 2001 at the University of Bari (Italy). He was then awarded a grant, funded by GlaxoSmithKline, to run a 2-year project entitled "Evolutionary Computation Methods for Drug Discovery" at the University of Sheffield (UK). Afterwards, Dr. Nicolotti was appointed as lecturer at the University of Bari (Italy). His interests are in the rational molecular design of inhibitors of therapeutical interest as well as in the derivation of QSAR predictive models according to REACH regulation. He authored 58 publications in international peer-reviewed journals and held 18 lectures at international meetings.

Marjana Novič was born in Ljubljana, Slovenia, where she took studies in Physical Chemistry at the University of Ljubljana during 1974-79. She obtained PhD in 1985 at the Faculty of Chemistry and Chemical Technology, University of Ljubljana, with the thesis "Hierarchical Clustering and Recognition of Chemical Structures and Structural Fragments on the Basis of 13C NMR Spectra." She started her career at the National Institute of Chemistry in Ljubljana, initially developing automated information systems in for infrared and NMR spectroscopy. Post-doctoral specialization was carried out at University of Lausanne, Switzerland (1986-87) in the field of automated pattern recognition in 2-D NMR spectra. In 1989, she obtained "Boris Kidrič" national award for scientific achievements. Later, she visited several research laboratories (in Tarragona and Cordoba, Spain, and in Hobart, Tasmania), while being employed in the Laboratory of Chemometrics at the National Institute of Chemistry, Ljubljana, Slovenia; currently,

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

she has a position of a head of the laboratory. Her expertise includes the development of chemometrics methods, QSAR and ANN modelling, structural elucidation of transmembrane segments of membrane proteins, innovative merging of chemometrics methods with molecular modeling, which facilitates an effective drug design. She is also teaching chemometrics at the University of Ljubljana.

Sudip Pan received his BSc (Chemistry Honors) and MSc (Chemistry) degrees from the Vidyasagar University, West Bengal. He joined the research group of Professor P. K. Chattaraj for his PhD degree in the Department of Chemistry, Indian Institute of Technology, Kharagpur, as a CSIR (Government of India) Fellow. At present, he is working as a Senior Research Fellow.

Tomasz Puzyn, PhD, DSc, works as Associate Professor at the University of Gdansk, Poland. He leads the Laboratory of Environmental Chemometrics at Faculty of Chemistry. Current projects of the group are focused on the development of computational methods for designing new products (nanomaterials, ionic liquids) that are safe for human health and the environment. Prof. Puzyn is the author of more than 50 research publications, including those in leading journals (e.g., *Nature Nanotechnology*). His works have been cited about 800 times and his Hirsch index is 18. Prof. Puzyn is an active member of the international nanosafety community: leader of two EU FP7 projects (NanoBRIDGES and NanoPUZZLES) conducted within the European NanoSafety Cluster and a member of the MODENA Action Steering Committee in frame of COST – European Cooperation in Science and Technology.

Bakhtiyor Rasulev is a QSAR group leader in Interdisciplinary Center for Nanotoxicity at Jackson State University. He received his PhD degree in Chemistry from the Institute of the Chemistry of Plant substances, Tashkent, Uzbekistan, in 2002. He joined the group of Prof. Jerzy Leszczynski in 2004 as a research associate. Dr. Rasulev researches a range of topics in structure-activity relationship studies, dealing with biological activity prediction of natural compounds, organic compounds, physico-chemical and toxicity prediction of various chemicals, including nanoparticles (metal oxides and carbon nanoparticles). For this purpose, he develops computational methods for assessing the physico-chemical properties, biological activity, toxicity and environmental risk of studied chemicals. He is the author of many contributions devoted to QSAR modeling and quantum-chemical applications. Dr. Rasulev has close collaboration with the Instituto di Ricerche Farmacologiche Mario Negri (Milano, Italy), University of Zagreb (Zagreb, Croatia), Johns Hopkins University (USA), etc. His accomplishments have been widely recognized. He is permanent reviewer of more than 17 peer-reviewed journals. Dr. Rasulev has received many scholarships and awards, including Scholarship of Drew University for participating in Residential School of Medicinal Chemistry (Madison, NJ), Young Investigators Travel Award from Toxicological Division of American Chemical Society, Travel Grant of CRDF Foundation, scholarship for short-term visit of the Institute of Desert study of Ben-Gurion University (Israel).

Achintya Saha, a faculty member of Dept. of Chemical Technology, University of Calcutta, has supervised nine PhD students and guided six research projects funded by different organizations. He has published 80 papers in standard journals. He was awarded SAARC fellowship in 2013.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Babak Sokouti has over 18 years IT technical management and consulting experience, including managing and maintaining sophisticated network infrastructures. He has obtained Bachelor of Science in Electrical Engineering with a specialization in Control from Isfahan University of Technology, Isfahan, Iran; a Master of Science in Electrical Engineering with a specialization in Electronics (with background of Biomedical Engineering) from Tabriz Branch, Islamic Azad University, Tabriz, Iran; a Master of Science in Information Security with Distinction from Royal Holloway University of London, London, UK; and is currently a PhD student in Structural Bioinformatics at Biotechnology Research Center, Tabriz University of Medical Sciences, Tabriz, Iran. In addition, he has obtained IT industry certifications including MCP, MCSA 2003, MCDBA 2000, MCSE 2003, and MCTS 2008. His research interests include cryptographic algorithms, information security, network security and protocols, image processing, protein structure prediction, and hybrid intelligent neural network systems based on genetic algorithms.

Alejandro Speck-Planche graduated in Chemistry in 2005, Department of Chemistry, Faculty of Natural Sciences, University of Oriente, Cuba. He is currently working as FCT researcher (2012-present) under the supervision of Prof. Dr. Maria Natália Dias Soeiro Cordeiro, at REQUIMTE/Department of Chemistry and Biochemistry, University of Porto. His main research interests are focused on the applications of chem-bioinformatics approaches in Drug Discovery, Nanotoxicology, Nanomedicine, Ecotoxicology, and Biophysics. Other fields of interest include the design of green pesticides, and the application of complex network theory for the study of multiple phenomena in Chemistry, Physics, Biology, and Environmental Sciences.

Venkatesan Subramanian obtained his Bachelor and Master of Science in Physics from the University of Madras in 1982 and 1984 and PhD in Physics from the University of Madras in 1993 under the supervision of Dr. T. Ramasami. He joined Central Leather Research Institute and heading the group of Computational Chemistry and Biology. His research interest includes non-covalent interaction in molecular cluster, quantitative structure activity relationship, collagen structure and stability, nanomaterial biomolecules interaction and energy materials (Solar and Hydrogen energy). He has published around 220 scientific publications in international journals. He is the fellow of The National Academy of Sciences, Allahabad, and Indian Academy of Sciences, Bangalore.

Khac-Minh Thai was born in 1977 in Hau Nghia Town, Long An Province, South of Vietnam. Thai graduated in Pharmacy in 2000 from the University of Medicine and Pharmacy at Ho Chi Minh City, Vietnam, and earned the Master's degree in Pharmacy in 2004 with Prof. Hyun-Ju Park from College of Pharmacy, Sungkyunkwan University, Seoul, Korea. He got his PhD degree in 2008 under the guidance of Prof. Dr. Gerhard F. Ecker at Emerging Field Pharmacoinformatics Group, Department of Medicinal Chemistry, University of Vienna, Vienna, Austria. Currently, he serves as Associate Professor at Department of Medicinal Chemistry, Faculty of Pharmacy, University of Medicine and Pharmacy at Ho Chi Minh City; 41 Dinh Tien Hoang Street, District 1, Ho Chi Minh City, Vietnam.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Alla Toropova has been working in Istituto di Ricerche Farmacologiche Mario Negri, Milano, Italy, with Dr. Benfenati's research group. She had published of about 120 papers in international journals on QSPR/QSAR analysis for inorganic and organometallic compounds, in particular thermodynamic properties of coordination compounds of bioactive metals and thermodynamic properties of nano substances. She has been working since over 15 years in the fields of Chemoinformatics and Nanoinformatics. She also has made scientific collaborations with many experts in the fields of Chemoinformatics and Nanoinformatics throughout the world. Dr. Toropova had taken part in development of approach of Chemoinformatics modeling based on optimal descriptors. The approach based on optimal descriptors can be useful alternative of traditional QSAR for the analyses of Nanomaterials as well as for Drug Discovery. She is one of developers of the software CORAL (http://www.insilico.eu/coral). She is participant in EC projects: ORCHESTRA, ANTARES, OSIRIS, NanoBRIDGES, PreNanoTox, NanoPUZZLES.

Andrey Toropov has been working in Istituto di Ricerche Farmacologiche Mario Milano, Italy, with Dr. Benfenati's research group. The optimal (flexible) descriptors are his main scientific interest. He is developer of recent version of the optimal descriptors named as collectors of eclectic data for QSPR/QSAR predictions. This paradigm is especially appropriate for the cases of complex molecular systems such as nanomaterials and proteins. He had done contribution for the development of the CORAL freeware (http://www.insilico.eu/coral). He had published of about 200 papers in international journals on the QSPR/QSAR analysis: physicochemical properties, toxicity, cancirogenicity, anticancer agents, anti-HIV-1 agents, QSPR/QSAR analysis for nanomaterials. He was participant of the following EC projects: DEMETRA, CHEMPREDICT, PROSIL, ORCHESTRA, ANTARES, OSIRIS, NanoBRIDGES, PreNanoTox, NanoPUZZLES, etc. He is a participant in work of Editorial Boards of scientific journals (as a reviewer). He was a visiting scientist at Jackson State University (USA, Mississippi), in 2006-2007.

Thanh-Dao Tran graduated in 1988, University of Medicine and Pharmacy at Ho Chi Minh City, Vietnam. Master degree: 1997, University of Medicine and Pharmacy at Ho Chi Minh City, Vietnam. PhD degree: 2004, Kangwon National University, Korea. Recent career: Senior researcher and Lecturer, University of Medicine and Pharmacy at Ho Chi Minh City, Vietnam. Title: Assoc. professor at Department of Medicinal Chemistry, University of Medicine and Pharmacy at Ho Chi Minh City, Vietnam.

Aleksandar Veselinović was born 1978 in Nis, Serbia. He graduated at University of Niš, Faculty of Science and Mathematics, Department of Chemistry as valedictorian in 2002. During studies, Aleksandar was scholar by Ministry of Education, Republic of Serbia and Serbian Academy of Science and Arts. He defended Master of Science thesis at University of Niš, Faculty of Science and Mathematics, Department of Chemistry in 2009. He defended Doctoral dissertation, scientific area Physical Chemistry at University of Niš, Faculty of Science and Mathematics, Department of Chemistry in 2014. Aleksandar works at University of Nis, Faculty of Medicine, Department of Pharmacy as teaching assistant. He is involved as a researcher in scientific projects funded by Ministry of Education, Republic of Serbia. Aleksandar's scientific areas of interests include physical and theoretical chemistry, chemoinformatics, and computer-aided drug design.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Jovana Veselinović was born 1983 in Skoplje, Republic of Macedonia. She graduated at University of Niš, Faculty of Medicine, Department of Pharmacy as valedictorian in 2008. During studies, Jovana was scholar by Ministry of Education, Republic of Serbia. Currently, Jovana is PhD student at University of Niš, Faculty of Medicine, Department of Pharmacy, scientific area Toxicology. University board approved her Doctoral thesis entitled "Pharmacochemical Aspects of Selected 4-Phenyl Hydroxycoumarins' Activities – Integrated in Vitro and Computational Study." She is supported as a researcher in scientific projects funded by Ministry of Education, Republic of Serbia. Jovana's scientific areas of interests include pathway analyses for drug discovery and target-design of new drugs.

Eleni Vrontaki was born on Crete, Greece, in 1987, and received her Bachelor's Degree (BSc) in Chemistry from the National and Kapodistrian University of Athens (UoA), Greece, in 2010. Furthermore, she attained a Certificate of Enology from the inter-university program of the University of Athens and the Agricultural University of Athens. In 2012, she received a Master's Degree (MSc) in "Organic Synthesis and Applications in Chemical Industry" (Organic Chemistry) from the UoA under the direction of Prof. Thomas Mavromoustakos. Since June 2012, she is a PhD student in Organic Chemistry at the same university, and by November of that year is a researcher in NovaMechanics Ltd, Chemoinformatics & Bioinformatics Solutions, Nicosia, Cyprus. Her expertise and research interests focus on Molecular Modeling, 3D-QSAR, and NMR Spectroscopy.

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Index

%CCT 534
2D-Descriptor 83
3D-QSAR 18, 21-26, 44, 126-127, 228-229, 255-256, 272, 276, 327-329, 335, 337, 339, 346, 371, 378, 380-382, 418, 429, 431
6-Fluoroquinolones 280, 289, 316
3Rs 180, 210

A

AChE 75, 354-364, 367-369, 378-387, 399
AChE Enzyme 75, 360-361, 367, 379, 387, 399
AcrAB-TolC 240-242, 244, 251, 254-255
Aliphatic Compounds 76, 125, 129, 141, 158, 177
Antibacterial Agents 26, 316
Antioxidants 212-215, 220, 222, 225-226, 230, 232, 355
Applicability Domain 27, 180-183, 190, 193, 198, 203, 210, 272, 278, 293-294, 371, 475, 477, 492, 494-495, 505, 511-512, 515
Aromatase 400-403, 406-418, 420-429, 431-439, 441-447, 470
Artificial Neural Networks (ANN) 3, 17-18, 28, 87, 113, 213, 224-225, 272, 276-277, 293-294, 316, 372, 510, 515-521, 234, 542
atom colors 91-92, 107

B

Bacterial Efflux Pumps 238-239, 242, 244, 247, 251, 254-256, 268
Breast cancer 14, 400-401, 403-409, 414, 446-447, 470

C

Chemical Graphs 49, 83
Chemoinformatics 50, 356, 376, 548, 550
Chemometrics 220
Competitive Inhibitors 342, 358, 361, 402, 414, 470
Computer Aided-Drug Design (CADD) 217, 236
CoMSIA 16, 22, 52, 86, 228, 276, 327, 335-336, 339, 371, 381-382, 400, 420, 426, 431, 444, 446, 470
ConGENER software 478, 480, 496
Connectivity 5, 7, 9-10, 17, 49, 54, 76, 83-84, 87, 129, 219, 562
CORAL software 512, 545-546, 560-562, 566-569, 571, 573-576
Correlation Weight 563, 585
Coumarin derivatives 213, 222, 230-232, 420
Curcumin 213, 222, 228-229, 232, 236

D

Dimensionality Reduction 44
DNA Gyrase 269, 271, 283, 289-292, 295-300, 316
Docking Study 268, 283, 382, 417, 422-423, 425, 427, 431, 437, 441
Drug-Likeness 1, 28, 111, 249, 273, 282, 293, 316

E

Ecotoxicity 155, 472, 494, 507, 509, 513-517, 520-523
Efflux Pump Inhibitors (EPIs) 238-239, 268
Electrophilicity 124, 129-130, 132-133, 136-137, 144, 146, 149, 151-152, 154-156, 158, 177, 226
Energies of Frontier Molecular Orbitals 177
Engineered Nanomaterials 537, 559
Environmental risks 507
EPI Suite 492, 505
Escherichia coli 240, 242, 251, 256, 511-512, 541, 545
Extended Topochemical Atom (ETA) Indices 48, 50, 53, 83

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

F

Flavonoids 213-215, 220, 222-226, 232, 237, 241, 249-250, 424-425, 444
Free Radicals 213-214, 226-227, 230, 237, 355

G

Global SMILES Attribute 585
Graph of Atomic Orbitals (GAO) 585
Group Philicity 124, 137, 155-156, 158, 177, 225

H

HCV Inhibitors 346, 353
Hepatitis C Virus 318-319, 327, 353
Hydrogen-Filled Graph (HFG) 585
Hydrogen-Suppressed Graph (HSG) 562, 585
Hydroxybenzalacetones 213, 222, 229, 232

I

In Silico Methods 232, 237, 269, 288, 356, 541
Integrated in silico methods 269
Interpolation 180-183, 196-197, 201, 203

L

ligand based drug design 403, 470
Local SMILES Attribute 585

M

Machine Learning 17-18, 28, 45, 195, 198, 224, 255-256, 341, 370, 417, 510, 512, 515, 534, 542
Mechanism-Based Inhibitors 402, 415, 470
MNPs 508-511, 534, 536, 541-543, 549-550, 552
Molecular docking 3, 109, 127, 256, 268, 272, 283, 286, 295, 327-328, 335-336, 356, 378, 381, 383, 386, 425
molecular dynamics simulation 333, 421, 426
Molecular Properties 7, 45, 75, 85, 128, 217, 226, 251, 254, 281, 326, 370, 470
Monte Carlo method 546, 549, 561, 564
Monte Carlo Optimization 566, 585
MultiDrug Resistance (MDR) 240, 268
Multimedia Mass-Balance Models 471, 493, 505
Multiple Linear Regression 3, 15, 87, 113, 200, 210, 213, 224, 226, 276, 381, 425, 434, 475, 510, 542, 545-546, 549, 571-572
multivariate image analysis (MIA) 84-85, 87-88, 91, 93, 109, 111, 113, 115, 122

N

Nanomaterials 493, 495-496, 506-511, 515, 523, 534-539, 541, 545, 547, 549-550, 552, 559, 574
Nanoobjects 536, 550, 559
Nanoparticles 493, 495, 506-508, 510, 515, 523, 534-541, 543-552, 559, 574
nano-QSAR 510-512, 514, 523, 534
Nanoscale 331, 507-508, 536-537, 539, 551-552, 559
Nanostructured Material 559
nanotechnology 506-510, 513, 535-536, 540, 549-552
neglected disease 84, 87
Net Atomic Charge 177, 341
Net Electrophilicity 124, 132-133, 137, 158, 177
non-steroidal aromatase inhibitors 400, 416, 429, 445
NorA 240-241, 244-247, 249-250, 255-256
NS3 318, 320, 334, 337-338, 342-343, 345-346
NS5B Polymerase 318, 326-327, 329, 332, 341, 353

O

OECD 27, 180-182, 278, 473, 491-492, 494, 496, 505, 511, 542, 550, 562, 575
Ontology 513, 516, 534
Optimal Descriptor 561-563, 566-567, 574, 585

P

Partial Least Squares (PLS) 3, 16, 23, 85, 87, 112, 210, 213, 276, 510, 548
partition coefficient 9, 17, 27, 51, 76, 86, 122, 130, 219, 483, 495, 569, 571
Persistence Organic Pollutants 505
Perturbation theory 513-514
Pharmaceuticals and Personal Care Products 493
Pharmacophore Hypothesis 231, 242, 244, 268
Pharmacophore Mapping 230, 232, 356, 386, 400, 417, 426, 431, 439, 441, 443-444, 446, 470
Pharmacophore modeling 3, 256, 272, 321-323, 421
Phenolic derivatives 213, 222, 229-230, 232
pixels 84-85, 88-91, 112
Polyaromatic Hydrocarbons 141, 158, 177
Polychlorinated Biphenyls 141, 177, 479
Principal Component Analysis 16, 21, 23, 122, 128, 210, 213, 220, 227, 251, 272, 432, 478, 489
Protein Data Bank 268, 321, 328, 337
Protein Homology Modeling 316
Protein sequence alignment 295
Protein X-Ray Crystallography 268

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use

Q

Quantitative Nanostructure-Activity Relationship (QNAR) 535-536, 541

R

Randomization 20, 211, 222, 224, 226, 230, 245, 340, 376
REACH Regulations 505
Risk Assessment 2, 180, 471, 474, 484, 492, 505, 509, 511-512, 514-515, 535-536, 538-539, 541, 543, 551, 559

S

Sensitivity of the Descriptors 534
Standard Deviation 14-15, 19, 194-195, 211, 327, 338, 376, 380, 429
Staphylococcus aureus 240, 244, 249, 295
structure based drug design 403

T

Test Data Set 20, 45, 384, 386
Tetrahymena Pyriformis 27, 129, 141, 177
Topology 5, 9, 24, 28, 48, 83, 126, 289-290, 371
Training Data Set 19, 45, 183, 198, 203, 224, 374-376, 384, 386
Trypanosomiasis 87, 93, 122
Tuberculosis 240, 269, 271, 283, 289, 291-292, 295-300

V

van der Waals radius 8
Virtual Combinatorial Library Design 270-271, 278-279, 281, 316
VmrA 251, 253-255

EBSCOhost - printed on 2/14/2023 7:16 AM via . All use subject to https://www.ebsco.com/terms-of-use