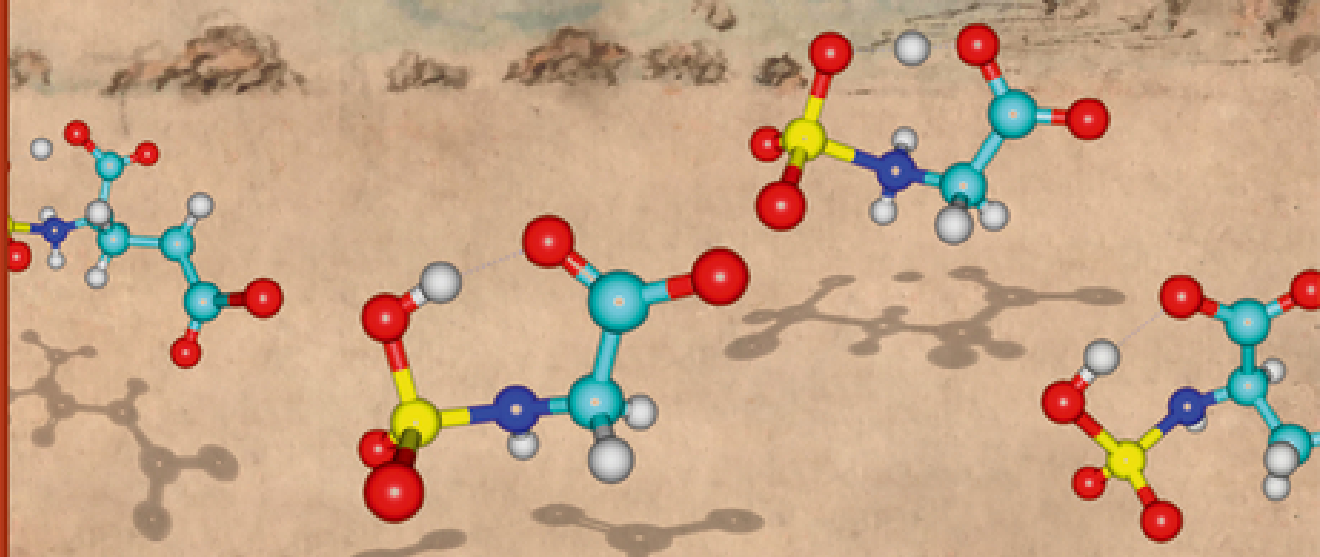
DE GRUYTER

Yufen Zhao, Yan Liu, Xiang Gao, Pengxiang Xu

PHOSPHORUS CHEMISTRY

THE ROLE OF PHOSPHORUS IN PREBIOTIC CHEMISTRY

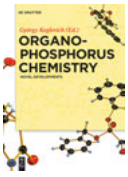


厦门大学出版社 XIAMEN UNIVERSITY PRESS 国家一级出版社 全国百佳图书出版单位

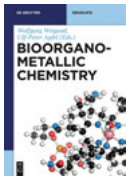
Yufen Zhao, Yan Liu, Xiang Gao, Pengxiang Xu

Phosphorus Chemistry

Also of Interest



Organophosphorus Chemistry.
Novel Developments
Keglevich (Ed.), 2018
ISBN 978-3-11-053453-5, e-ISBN 978-3-11-053583-9



Bioorganometallic Chemistry.
Weigand, Apfel (Eds.), 2019
ISBN 978-3-11-049650-5, e-ISBN 978-3-11-049657-4

Biotechnological Applications of Extremophilic Microorganisms.
Lee (Ed.), 2018
ISBN 978-3-11-042773-8, e-ISBN 978-3-11-042433-1

Molecular Approaches to Evolution.
Ninio, 2014
e-ISBN 978-1-4008-5624-4

Yufen Zhao, Yan Liu, Xiang Gao, Pengxiang Xu

Phosphorus Chemistry

The Role of Phosphorus in Prebiotic Chemistry

DE GRUYTER

厦门大学出版社 XIAMEN UNIVERSITY PRESS 国家一级出版社 全国百佳图书出版单位

Authors
Prof. Dr. Yufen Zhao
Xiamen University
College of Chemistry and
Chemical Engineering
422 Si Ming South road
Xiamen
PR China
yfzhao@xmu.edu.cn

Dr. Xiang Gao
Xiamen University
School of Pharmaceutical
Sciences
Xiang An South road
Xiamen
PR China
xgao@xmu.edu.cn

Dr. Yan Liu
Xiamen University
College of Chemistry and
Chemical Engineering
422 Si Ming South road
Xiamen
PR China
stacyliu@xmu.edu.cn

Dr. Pengxiang Xu
Xiamen University
College of Chemistry and
Chemical Engineering
422 Si Ming South road
Xiamen
PR China
xpengxiang@xmu.edu.cn

ISBN 978-3-11-056237-8
e-ISBN (PDF) 978-3-11-056255-2
e-ISBN (EPUB) 978-3-11-056245-3

Library of Congress Control Number: 2018950586

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at http://dnb.dnb.de.

Typesetting: Integra Software Services Pvt. Ltd.
Printing and binding: CPI books GmbH, Leck
Cover image: courtesy of the authors

www.degruyter.com

Foreword

The origin and evolution of life under prebiotic conditions has been one of the frontier issues that scientists across the world endeavor to solve. Charles Darwin, English biologist and founder of the evolution theory, has put forward a bold hypothesis at the end of his masterpiece *On the Origin of Species*: "In some warm little ponds with ammonia and phosphates, proteins will be formed under lighting and heating conditions." He believes that phosphorus and nitrogen are a crucial part during the origin of life. Phosphorus plays a leading role in life on Earth. About 9% of our DNA is made of phosphorus. The half-life of DNA's spontaneous hydrolysis can be as long as 31 million years. Phosphate esters are quite stable; thus, phosphorus cannot be replaced by other elements such as silicon, arsenic and sulfur. Just as the Noble Prize winner Professor L. Todd said, "Where there is life, there is phosphorus."

In 2013, scientists found that the content of phosphates on Mars was 5–10 times of that on Earth. So here's the question. Do these phosphates come from life? Can phosphorus be the indication for exploring extraterrestrial life? And why nature chooses phosphorus? This book is a collection of the research achievements of our group during the last decades. The activation mechanism of penta-coordinated phosphorus, which is dominated by phosphorus chemistry, plays a vital role in the regulation of life system. We can say that α-amino acids are chosen by phosphorus. We have proposed that the *N*-phosphoryl amino acid is the cogenetic evolutionary model of the smallest molecule for nucleic acids, proteins and membranes. We have found that Ser–His is the smallest hydrolase of proteins and nucleic acids. We have put forward the research model for chirality origin and coden origin. Our research findings provide scientific basis for human beings to unveil the origin of life.

We are honored that Professor Gerda Horneck, Honorary President of the European Astrobiology Network Association, and Professor Jisheng Chen, Academician of CAE at the Fourth Research Institute, People's Liberation Army Institute of Chemical Defense, wrote a preface to the book. Many people have participated in writing this work. They are Prof. G. M. Blackburn and Dr Yi Jin from Sheffield University, Dr Feng Ni from the University of Southern California, Prof. Shuxia Cao and Dr Jihong Liu from Zhengzhou University, Prof. Hongyu Zhang from Huazhong Agricultural University, Prof. Yanmei Li and Dr Haiyan Wang from Tsinghua University, Associate Professors Yan Liu, Xiang Gao, Pengxiang Xu, Daxiong Han and Zhigang Chen from Xiamen University, Dr Zhiping Zeng from Xiamen University, as well as PhD candidates including Yile Wu, Yongfei Yu and Liu Liu from Xiamen University. I thank them for their hard work. I am grateful to Mr Wei Sui, associate senior editor in Xiamen University Press, for planning and editing the book. My gratitude to Xiamen University Press for the great support for the publication of this book. I also appreciate

https://doi.org/10.1515/9783110562552-201

the financial support of the publishing fund for the Sixth Nanqiang Book Series of Xiamen University.

March 2016

Yufen Zhao

Preface I

Prebiotic chemistry of phosphorus chemistry and origin of biomolecules

It is a pleasure and honor for me to provide the preface for the book *Phosphorus Chemistry: The Role of Phosphorus in Prebiotic Chemistry* edited by Professor Yufen Zhao on the occasion of the 95th anniversary celebration of Xiamen University. There is no other group in the world better qualified to review the role of phosphorus in the prebiotic formation of essential biomolecules, such as amino acids, nucleosides, nucleotides, oligonucleotides and finally peptides as well as nucleic acids. This review is based on nearly 30 years of systematic research by Professor Zhao and her group on phosphorus and life, having identified phosphorylation as one of the key events in prebiotic evolution on the early Earth. *N*-phosphoryl amino acids play an essential role in this scenario. It was found that α-amino acids after N-phosphorylation could conduct a series of reactions important for the origin of life, such as peptide formation, ester exchanges, phosphoryl group migration, nucleotide formation, as well as membrane formation. It was demonstrated that only the α-amino acids could be activated to form a peptide, while the β- and γ-amino acid remained unchanged under the same conditions. Phosphoryl amino acid 5′-nucleosides having a P–N bond have been described as a model of the origin of amino acid homochirality and the genetic code: the chiral selection of the earliest amino acids for l-enantiomers seemed to be determined by a stereochemical/physicochemical relationship. On the basis of this abundance of data on the role of phosphorylation in prebiotic processes, a comprehensive model of coevolution of proteins and nucleic acids in the origin of life scenario was formulated. The 11 chapters of this book provide an insight into this work on the importance of phosphorus for life, in prebiotic chemistry as well as in contemporary life.

Gerda Horneck

Former Deputy Director of the Institute of Aerospace Medicine, German Aerospace Center DLR, Cologne, Germany (retired)
Honorary President of the European Astrobiology Network Association EANA

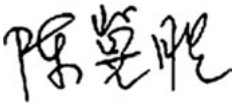
https://doi.org/10.1515/9783110562552-202

Preface II

The origin of life has always been a mystery to human beings. Essentially, it complies with Charles Darwin's theory of evolution, and its process should also be a chemical evolutionary process from simple to complex. The prebiotic synthesis of basic substances of life, how micromolecules evolve into macro ones with biological functions, energy metabolism, as well as inheritance and duplication of information, all these are the research focuses of the origin of life. In 1953, American scientist S. Miller conducted the famous spark-discharge experiment. He synthesized a variety of amino acids based on micromolecules, thus opening a new era for the chemical evolution of life. During the past 60 years, many breakthroughs have been made in the study of the origin of life. This book is filled with the research achievements of Professor Yufen Zhao for over 30 years at the laboratories of the Institute of Chemistry, Chinese Academy of Sciences, Tsinghua University, Xiamen University and Zhengzhou University. This work features analysing some key issues concerning the origin of life from the basics of phosphorus chemistry. These achievements detailedly demonstrate the catalysis and regulation of phosphorus for living substances under prebiotic chemical conditions to readers, providing novel ideas and clues for solving the mystery of the world, namely, the origin of life.

24 March 2016

Jisheng Chen



https://doi.org/10.1515/9783110562552-203

Contents

1 The international background of the origin of life

Where did life come from? The "extraterrestrial life hypothesis" postulates that the seeds of life originated from the outer space. However, life might have possibly originated on the Earth based on Miller's reaction data. At the same time, phosphorus played a key role in the production of vital essential substances such as amino acids, proteins, nucleosides and nucleotides.

The origin of life is one of the eternal mysteries of mankind. Since Miller performed the discharge reaction in 1953, many scientists have performed countless experiments on the origin of life. They utilized inorganic small molecules containing C, H, N, O, S and P elements to build life materials: amino acids, nitrogenous bases (adenine (A)/guanine (G)/cytosine (C)/uracil (U) and nucleosides. They were then oligomerized into oligopeptides and oligonucleotides. These two kinds of building blocks are the basis of "protein origin" and "RNA origin." In addition, the formation of the membrane is one of the most necessary conditions for the protocells. Therefore, a more reasonable original living system should contain proteins, nucleic acids and membranes. In 2015, Sutherland proposed that HCN, H_2S, acetylene and phosphorus could simultaneously produce amino acids, uracil monophosphate (UMP) and phospholipid predecessors via a copper oxide oxidation reaction [1]. However, even if this co-origin hypothesis is reasonable, the most fundamental problem of primitive life should be the origin of genetic code. The origin of life could be in an order way only under the guidance of an efficient codon blueprint. The genetic material of life on Earth is DNA and phosphorus accounts for 9% of DNA. Phosphate diester plays a unique function in life activities. However, where the phosphorus comes from on the Earth? And is there presence of phosphorus on other planets? [2].

Although phosphorus is one of the trace elements in the universe, it is present at a much higher level in our life. Phosphorus in our life plays many basic biochemical functions, such as information storage by nucleic acids, energy transfer by adenine and guanine nucleotides, phospholipids and signal transmission by cyclic nucleotides.

1.1 Original source of phosphorus

1.1.1 Nuclear reactions

For nuclear reactions, in the synthesis of H, C, O, N, S and P (Table 1.1), the synthesis of phosphorus is quite different. First, the generation of phosphorus nuclei is carried out from the fuel of C and Ne only on a few celestial bodies with sufficient mass. Second, the yield of phosphorus is low (the combined yield is 2.5%) in many phosphorus nuclei-producing reactions. This fact also explains why phosphorus content is relatively low in the main elements of life.

https://doi.org/10.1515/9783110562552-001

Table 1.1: Nuclear reactions of life-related elements.

	Source	Process	Rank
H^{1}	Big-Bang	Primordial nucleosynthesis	1
He^{2}	Big-Bang and H-burning	Primordial synthesis p-p thermonuclear reaction and CNO cycle	2
O^{8}	He-burning	Alpha-capture: $^{12}C + \alpha \Longrightarrow {}^{16}O$	3
C^{6}	He-burning	Triple-alpha process: $3\alpha \Longrightarrow {}^{12}C$	4
Ne^{10}	C-burning	Alpha-capture: $^{16}O + \alpha \Longrightarrow {}^{20}Ne$	5
N^{7}	H-burning	CNO cycle	6
S^{16}	O-burning	$^{16}O + {}^{16}O \Longrightarrow {}^{28}Si$ $^{28}Si + \alpha \Longrightarrow {}^{32}S$	10
P^{15}	C- and Ne-burning	Many processes	17

1.1.2 Phosphorus surrounds celestial interval and interstellar space

There is little knowledge about the chemical properties of phosphorus that surround celestial bodies and interstellar spaces. Nonetheless, phosphorus has an unusual position compared to other major elements of life. The abundance of phosphorus is relatively rich in cosmic space and is at 17th among all elements in the universe [3]. At present, only phosphatides [4] and organophosphorus compounds [5] have been detected in interstellar space. In addition, neither PH_3 nor phosphorus oxides may exist on the primitive Earth [6]. However, phosphorus oxides were predicted as the main form of phosphorus in interstellar space from the viewpoint of energy. This might be because of the observation error, as oxygen is mainly present in the form of carbon oxides in the detection of carbon-rich celestial bodies.

1.1.3 Phosphorus in meteorites

The measurement of the composition of pellet meteorites, stone iron meteorites, lunar samples and Martian meteorites indicates that phosphorus is scarce but widely present in the solar system. Interestingly, the content of phosphorus in Shergotty, Nakhla and Chassigny (SNC) meteorites is higher than other types of meteorites and Earth rocks. Although these samples cannot represent all stars, it is very meaningful to explore the task of Mars in the future.

The analysis of the elements of Mars generally depends on the analysis of SNC series meteorites. The results are shown in Table 1.2.

In 2013, the spirit Mars probe in the United States detected phosphorus on Mars to be 5 to 10 times more than that present on the Earth [2].

Table 1.2: Meteorite phosphorus content.

	Shergotty	EETA 79001A	EETA 79001B	ALHA 77005	Nakhla	Chassigny
P_2O_5 (wt%)	0.80	0.54	1.31	0.36	0.10	0.06

1.1.4 Phosphorus in comets

It was found that comets may contain certain amounts of phosphorus by the thorough monitoring of phosphorus anions PO_2^- and PO_3^- and phosphate rocks in interstellar dusts; these amounts are at least comparable to that found in the Sun. Thus, the evidence of presence of phosphorus oxides and phosphates in the center of the comet dust particles is still a challenging task, even though the present form of phosphorus in the comet is not clear yet.

1.1.5 Alkyl phosphoric acid homologues

By studies conducted on alkyl phosphoric acid derivatives, it is found that organophosphorus compounds involved in pre-biochemical processes are transported primarily to the original Earth by the impact of stars and meteorites. Similarly, most of water and volatile matters present on Earth are converged by comets.

1.1.6 Phosphorus in the crust

The total mass of the existing phosphorus in the crust is about 8×10^{20} g; it is 11th among all the elements present. Phosphorus content is rich in volcanic rocks, deformation rocks and sedimentary rocks. Scientists have calculated that comets and meteorites captured by primitive Earth at the later stage of growth may contribute to 10% of the total mass of phosphorus. It may provide a hypothetical view of phosphorus-containing compound to the original transport process of the Earth. Although phosphorus may be present in both positive trivalent and pentavalent states, almost all the phosphorus present in the lithosphere of the Earth is the phosphate type PO_4^{3-}.

Table 1.3 below shows the presence of phosphorus on Earth. It is worth mentioning that the terrestrial rocks and seabed sedimentary rocks have similar levels of phosphorus, but less phosphorus content in the oceans. Although it is not explicitly stated in this table, we predict (we predict the quality of C is 10^{18} g) that the total amount of phosphorus captured in the Earth's biosphere is about 4×10^{16} g.

Table 1.3: Phosphorus on the Earth.

Reservoir	Mass content (g)
Igneous and metamorphic rocks on land	4.3×10^{20}
Marine sedimentary rocks	3.9×10^{20}
Oceans	6.3×10^{16}
Total content	8.2×10^{20}

1.1.7 Phosphorylation under prebiological conditions

In 1964, Miller and Parris succeeded in simulating pre-biophosphorylation at a moderate temperature (35°C). They obtained 27% pyrophosphates with hydroxyapatite and potassium cyanate. At the same time, a small amount of triphosphate (about 5% of pyrophosphate) was obtained. According to their results, cyanate and phosphate can be properly synthesized on the surface of the hydroxyapatite, and the product can continue to react with other phosphates to generate pyrophosphate.

Combined with other experiments, it can be considered that hydroxyapatites, phosphates and ammonium salts may all be involved in the local environment of the phosphorylation reaction. As Charles Darwin predicted in the letter to Hooker, "all the ammonium salts and phosphates are unusual." With this point in hand, if the nucleotides are successfully phosphorylated on the surface of the evaporated ponds and ores, the nucleic acids can be condensed into considerable length of nucleotides on the surface of the clay and other ores, and carry out the original molecular copy.

Schwartz mentioned that most of the natural inorganic phosphorus in the Earth and meteorites was not water soluble at pH 8, but if the pH drops from 5 to 6, the solubility could be greatly improved [7]. They believed that phosphites could be used as a biosynthetic phosphorylating agent.

1.1.8 Outlook

In order to verify whether the phosphorus-containing compounds on the original Earth are carried by meteorites and comets in the prebiological processes, the following studies can be carried out:

1. Detection of specific phosphorus-containing groups in interstellar space, mainly in the interstellar space of oxygen-enriched objects in PO and PO_2 and HCnP or HPnN in the giant dense molecular cloud.
2. Equip with high performance mass spectrometry equipment in the detection of future comet spacecraft to simultaneously monitor phosphorus oxides and organic phosphides [8].

3. Accurate and detailed analysis of the elements of the original comet samples, such as elemental analysis, mass spectrometry and mineralogical analysis. It is possible to confirm the presence of such phosphates or pyrophosphate-containing particles during the prebiological period by analyzing the results.
4. Efforts to establish a prebiological phosphorylation evolution model [9–10] from the theoretical point of view. Focus on the catalytic reactivity of phosphates and other active phosphorus compounds in the primitive life.

1.2 Phosphorus and small biomolecules

1.2.1 Formation of amino acids with phosphorus

On account of the detection of PH_3 in the atmosphere of Jupiter in 1974, [11] it was considered as one of the original atmospheric mixture components. Based on this data, it was suggested that PH_3 might be the source of phosphorus necessary for the synthesis of basic molecules and played an important role in the evolution of primitive Earth. In 1983, Wang Wenqing and Kobayashi Kensei carried out the pre-biosynthesis of amino acids in a mixed gas consisting of CH_4, N_2, NH_3, H_2O and PH_3 in the Ponnamperuma Cyril lab [12–14]. Compared with the control system without PH_3, gas chromatography showed that the PH_3-containing system produced 19 kinds of amino acids after the discharge, but PH_3-free system under the same discharge conditions produced only six amino acids, including alanine, glycine, and aspartic acid (Table 1.4). In this reaction, almost all the protein-forming amino acids were generated. The result of this experiment showed that phosphorus might play an important role in the formation of amino acids in the early stages of organic synthesis in the origin of life. Thus, PH_3 may be a catalyst in the gas phase and aqueous phase during chemical evolution.

1.2.2 Formation of bases

Purine and pyrimidine are important bases to make up the nucleic acid. These compounds have been synthesized by complex organic chemical methods (abiotic pathways).

Fox and Harada mixed malic acid, urea and trimetaphosphate (P_3m); when heated at 100°C–140°C, uracil was obtained [15]. The amount of polyphosphate increased the yield of uracil. In this work, for the first time it was postulated that the polyphosphate reagent might have existed in the original Earth , and its impact on the formation of nucleoside base was studied.

In 2015, the Czech scientist Svatopluk Civiš simulated asteroid impact on the Earth conditions about 4 billion years ago, and treated formamide and soil by plasma.

Table 1.4: $PH_3 + N_2 + CH_4 + NH_3 + H_2O$ system after cremation discharge of amino acid production.

Peak	Amino acid	P-1 yield/nmol	P-2 yield/nmol	N-1 yield/nmol
1	Sarcosine	230	2,700	–
2	Alanine	57	540	18,000
3	α-Alanine	75	–	–
4	Valine	44	–	34,000
5	Glycine	970	6,700	5,800
6	β-APA	6	120	20
7	Isoleucine	73	280	–
8	Leucine	59	55	–
9	*N*-leucine	100	140	–
10	Proline	32	110	–
11	Threonine	110	79	–
12	Serine	1,000	63	–
15	Aspartic acid	160	280	11,000
17	Phenylalanine	46	160	–
19	Glutamic acid	250	260	100
22	a,β-Diaminopropionic acid	74	44	–
24	Tyrosine	5	–	–
26	Ornithine	610	250	–
27	Lysine	31	110	–

Note: P-1, P-2 gas phase are CH_4 and N_2in 26.7 kPa pressure; PH_3 is (4.0–9.33) kPa; 100 mL water, NH_4^+ 50 mmol / L, pH 8.0–8.7; temperature 60°C, cremation discharge at 24 h; N-1 gas phase does not contain PH_3; the other conditions are the same.

A (79.4%), G (4.4%), C (13.1%) and U (3.1%) could be produced in the reaction at the same time [16]. As shown in Figure 1.1, they proposed a free radical mechanism, which for the first time in the world showed the detection of four nucleoside bases simultaneously in a whole reaction system.

1.2.3 Formation of nucleosides with phosphorus

It is a key step for the combination of ribose and base to form nucleosides in the production of vital substances of vital origin.

In 1963, Ponnamperuma made an attempt to utilize UV to form nucleosides [17]. A dilute solution of ^{14}C-labeled adenine and ribose was mixed with phosphoric acid or ethyl orthophosphate under ultraviolet light (253.7 nm) and adenosine was obtained after the reaction. It was worth noting that adenosine cannot be obtained without phosphoric acid. Thus, the formation of adenosine needs to be in the presence of phosphoric acid or phosphate. In 1964, Ponnamperuma and Kirk reported that deoxyadenosine could be synthesized from a dilute solution of adenine

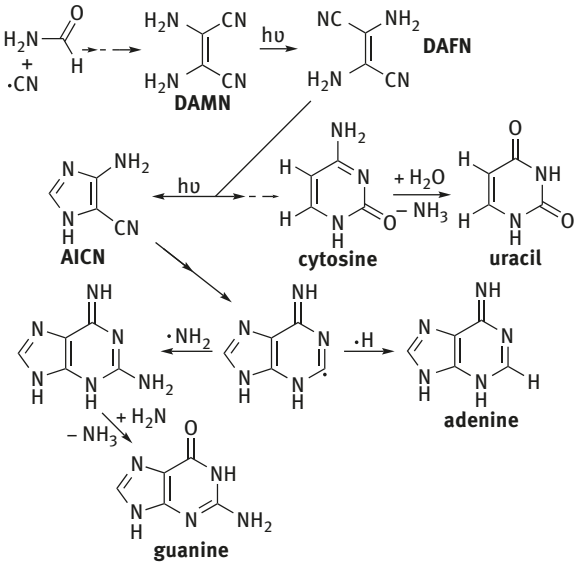


Figure 1.1: The conversion mechanism of A/G/C/U from formamide treated with plasma.

and deoxyribose under the light condition [18]. Similarly, deoxyadenosine cannot be obtained in the absence of phosphoric acid. Deoxyadenosine was produced at a yield of 5% by the action of ultraviolet light (253.7 nm) in the presence of phosphoric acid. In 1962, Schramm used polyphosphoric acid as a condensation agent. Adenine and deoxyribose condensed into deoxyribonucleoside in the dimethylformamide solution [19]. In 1963, Carbon studied the activity of polyphosphoric acid in the synthesis of nucleosides in different systems [20]. In an organic system, the yield was more than 50%, but the abovementioned synthetic yield was extremely low in aqueous solution system.

1.2.4 Formation of nucleotides

Riboses and deoxyriboses are able to convert into nucleotides via phosphorylation with nucleosides. Ponnamperuma and Mack (1965) found the phosphorylation of nucleosides occurred with a mixture of adenosine, guanosine, cytidine, uridine, thymidine and various phosphates at 160°C [21]. In 1957, Beck [22] found uridine and inorganic phosphate converted into uridine 5′-phosphate (5′ UMP), 2′(3′) -uridinephosphate and uridinediphosphate at 65°C–85°C for a long time. In 1968, Lohrmann and Orgel [23] found phosphorylation of uridine via a condensation reaction with cyanide, hydrocyanic acid, cyanamide, dicyandiamide and water-soluble carbodiimide in the presence of $H_2PO_4^-$, and UMP was obtained with 1% to 4% yield.

Schwartz and Ponnamperuma (1968) heated adenosine with linear polyphosphate in aqueous solution to obtain 2′-, 3′- and 5′-adenosine monophosphate [24]. As

this polyphosphate was able to phosphorylate a variety of nucleosides over a wide pH range, it might be one of the phosphorylating agents on the primitive Earth. Weanheldt and Fox (1967) found polyphosphoric acid could phosphorylate adenosine, guanosine, uridine, cytidine, deoxycytidine and thymidine at a temperature of 0°C to 22°C with a total yield of 25% to 45% [25].

Many scientists believe that the most abundant phosphorus-containing minerals on the original Earth are difficult to act as good phosphorylating agents due to insolubility and low reactivity. However, Miller and Parris found that polyphosphoric acid was formed on apatite surface via a CN-heating reaction [26]. Thus, it was concluded that phosphorylation of nucleosides can be involved in the formation of polyphosphoric acid from insoluble and highly reactive phosphates. Soon afterward, this inference got evidence. Trimetaphosphate (P_3m) reacted with nucleosides for four days to obtain high yields of 2′- and 3′-nucleotides [27]. In 1991, Yamaganta demonstrated that volcanic activation could produce large amounts of polyphosphoric acid [28], and P_3m was able to be present on the Earth's surface under original conditions. Due to the high reactivity of P_3m, several strategies for the synthesis of P_3m from phosphoric acid were proposed. For example, the reaction of nitrile with phosphate produces Cyanovinyl Phosphate (CVP) ($NCCH{=}CHOPO_3^{2-}$), CVP and P_3m were heated in aqueous solution to obtain phosphates, [29] and synthesis strategy was obtained by heating phosphoric acid or monobasic phosphate under anhydrous conditions to obtain P_3m [30]. Etaix and Orgel deeply studied polyphosphoric acid and nucleoside reaction, and found that 10% to 40% of deoxynucleoside triphosphate was obtained when combining deoxynucleoside and P_3m at 25°C in alkaline condition. However, only a small amount of nucleoside triphosphate was obtained with ribonucleoside in the same conditions. Tsukako's later work proved that the ribonucleoside had much higher reactivity than the deoxyribonucleoside, and almost completely converted into 2′-, 3′-nucleotides [31–32]. The reaction initially formed a small amount of 2′, 3′-cyclic phosphoryl nucleoside, which was hydrolyzed to 2′-, 3′-nucleotide. Polyphosphoric acid probably existed on the original Earth and was able to phosphorylate nucleosides and sugars [33–34]. Thus, it was the most probable phosphorylating agent under primitive conditions. Rabinowitz's work in 1969 and Yamanaka's work in 1988 proved that P_3m was the most effective condensation and phosphorylation agent produced in all volcanic activity [35].

1.3 Phosphorus and biomacromolecules

1.3.1 Formation of peptides with phosphorus

Gibbs free energy is increased in the process of amino acid dehydration–condensation connection (Figure 1.2). Without the addition of energy, the reaction of producing

Figure 1.2: Amino acid dehydration and condensation to form peptides.

peptides does not proceed spontaneously. For the synthesis of peptides, in modern organic chemistry, the amino, carboxyl and active side chains of amino acids need to be protected, followed by the condensation reaction. The reaction is carried out by the carboxyl group activation. There is also a process of carboxylation activation in peptide biosynthesis. However, it is almost impossible to consciously protect amino and carboxyl groups in the original Earth conditions, and the ratio of this chance is quite small.

There may be some dehydration condensers on the original Earth. For example, polyphosphates have an ability to promote the formation of peptides in aqueous solutions. Fox added polyphosphates in the condensation reaction of amino acids. They heated phosphoric acid at 200°C, 250°C, 300°C or 350°C to generate polyphosphates, followed by the condensation of amino acids. The condensation reaction is generally carried out at 100°C, and occurred even at a temperature of 65°C [36]. The phosphoric acid obviously played the role of dehydrating agent, solvent and catalyst in this reaction. However, under the same conditions, concentrated H_2SO_4 cannot promote the condensation of amino acids. In this biological evolution point of view, the catalytic role of phosphorus is very important.

In the chemistry of thermodynamics, the presence of water was not conducive to the spontaneous progress of amino acid polymerization reaction, but was conducive to transport the reactants. Water also protected the products from damage from the Sun's radiation. Thus, it is more biologically significant to study amino acid condensation reaction in aqueous solution.

In 1969, Feldmann and Rabinowitz used polyphosphoric acid for amino acid condensation in water [37]. They found that glycine and α-alanine can get the ideal yield of dipeptides with the addition of P_3m. Rabinowitz extended this reaction to a reasonable prebiotic conditions (weakly alkaline, low temperature and low concentration), [38] and also obtained satisfactory results. The best reaction condition occurred at pH 7.5–9.5, temperature 70°C, and for 70 h, with the yield of dipeptide up to 35%. Based on the experimental results, Feldmann and Rabinowitz proposed a peptide pathway via an acyl phosphate intermediate [37, 39]. Carboxyl group of the amino acid attacked the phosphorus atom of P_3m to form this intermediate (Figure 1.3). In their earlier work, it was found that the phosphate carboxylic acid anhydride reacted with amino acids to form peptide bonds. This result also supported their speculation.

In 1976, Hulshof and Ponnamperuma believed that linear and cyclic polyphosphoric acid should be the most likely condensate on the original Earth, whether from a chemical point of view or from a biological point of view [40].

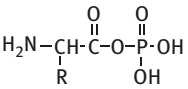


Figure 1.3: The mixed anhydride intermediate that Feldmann and Rabinowitz proposed in the peptide reaction with P_3m.

1.3.2 Nucleic acid pre-biosynthesis

Prebiotic synthesis refers to the natural synthesis of all possible abiotic processes under the conditions of the primitive Earth, before the organism appears. It is a major step from simple nucleotides to RNA in chemical evolution. In the absence of enzymes, issues of how to accumulate a certain amount of nucleotides and how the polymerization reaction is carried out are very important. In theory, nucleic acid pre-biosynthesis refers to the nucleotide polymerization, the first formation of oligonucleotides, and then further dehydration of oligonucleotides to form polynucleotides, and it ultimately evolved into RNA macromolecules. Depending on the simulation of primitive Earth conditions, the way of possible aggregation of nucleotides can be roughly divided into three categories till now. One is under the heating or light condition. Another one is in the aqueous solution and the presence of reasonable condensation agent. The third one is under a variety of possible template guidance.

(1) Heating or radiation-induced polycondensation of nucleotides: Schramm [41] found that nucleotides were polymerized at 50°C with ethyl metaphosphate, and the polynucleotides were obtained at about 20% yield. Schwartz and Fox (1967) [42] mixed 5′-phosphate cytidine and other nucleosides at 65°C, and produced oligonucleotides with an average degree of polymerization of 5.6. Over 50% to 60% of the oligonucleotides are present as 2′-,5′- and 3′-,5′-phosphodiester bonds. The key to this experiment is polyphosphoric acid, which is a good condensation catalyst for thermal polymerization. Although many scientists doubt whether primitive Earth can accumulate such amount of polyphosphoric acid, some others believed that the phosphoric acid in local areas (especially volcanic areas) has the potential ability to form polyphosphate due to the heat released by volcanic activity [35]. Moravek published a number of research articles on the thermal polymerization of uridine [30]. They heated 5′-phosphoric acid uridine to 160°C to give oligomers with substantially 2′-,5′- and 3′-,5′-sites. In 1962, Contreras [43] polymerized ribonucleotides with gamma rays irradiating in aqueous solutions.

(2) Condensation of nucleotides in the presence of condensation agents: This reaction usually uses imidazole and cyanamide as dehydration and condensation agent. Cyanamide-based condensers have been used to produce phosphate and peptide bonds of sugars in aqueous solutions, and it is probably a method of synthesizing oligonucleotides. According to the geological points, Steimann and Capolupo [44] suggested that the primitive Earth could generate these compounds. Since imidazole and cyanamide could be formed on the original Earth by abiotic pathways,

such reactions had very important prebiological significance. The temperature of these reactions was relatively low (60°C–100°C), and the reaction time was very short (18–40 h). Kaolin [45] and poly-L-ornithine were moderate catalytic agents, and usually form 2- to 10- mers polynucleotides [46].

(3) Condensation with nucleoside phosphorimidazolides: In the study of oligonucleotides produced by the polymerization of monomeric nucleotides, the most widely used reagent is imidazole-activated nucleotide derivatives, such as ImpA, (adenosine-5'-phosphorimidazolide), ImpG, ImpC and ImpU. Lohrmann discussed the pre-biosynthetic pathway of these derivatives [47]. ImpA might be formed by a mixture of adenosine 5′-polyphosphate or P1, P2-diadenosine diphosphate (A^5'ppA) with Mg^{2+} and imidazole at a suitable temperature. The energy was transferred from the original polyphosphate (p_nA, n≥3, such as ATP) to the imidazolyl phosphoric acid derivative. In addition, p_nA might be produced by the reaction of nucleotide 5′-AMP with P_3m (Figure 1.4).

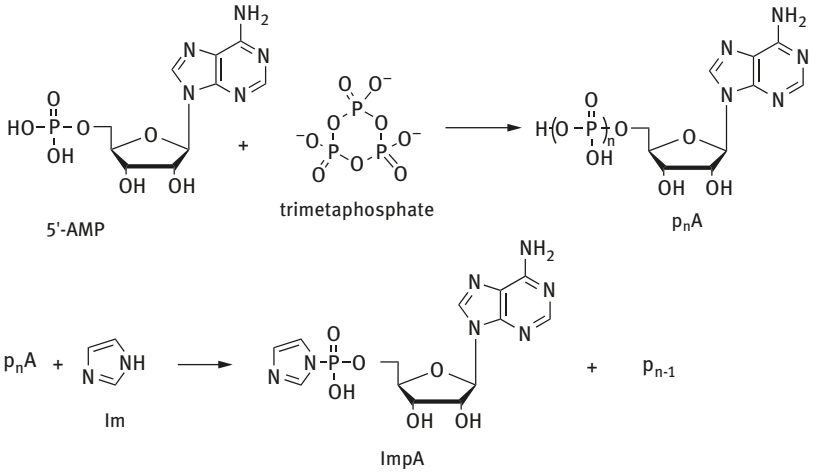


Figure 1.4: ImpA's pre-biosynthetic pathway.

A large number of experimental facts show that inorganic mineral clay can catalyze the formation of oligonucleotides. Mineral clay is widely distributed on the Earth, and it can catalyze a wide variety of chemical reactions. Most of these minerals perform the flaky structure of aluminosilicate as its main ingredient. As early as 50 years ago, it was suggested that inorganic minerals may be the possible prebiocatalyst on the Earth. It was even speculated that minerals were the basic form of primitive life [48]. These inorganic minerals, which are widely present on the original Earth, can absorb and concentrate the pre-biomolecules in the aqueous

solution so that these organic molecules are arranged on a solid surface with a tendency to form macromolecules among the various molecules. Ferris discovered that montmorillonite could act as an abiotic template to promote activation of nucleotide aggregates [49–51]. In an alkaline solution (pH = 8), ImpA produces 10-mers on montmorillonite, and forms a regional orientation of the 3′-,5′-phosphodiester bond by adding 10% diadenosine pyrophosphate to the reaction system. Without montmorillonite, however, the reactant ImpA is hydrolyzed in aqueous solution to form 5′-nucleotides and reacts with unhydrolyzed ImpA to form dinucleotide pyrophosphate [52]. The study of ImpC and ImpU under the same conditions shows that the abovementioned conclusions equally apply. For different activated groups, nucleotides activated by benzimidazoles and o-dimethyl aminopyridine derivatives give the highest polymer yield.

(4) Nucleotide synthesis under the template guidance: The templates for nonenzyme-based polymerization are mainly poly(U), poly (C), poly (U–G) and poly (A–C), which are believed to promote the polymerization reaction of complementary nucleotides or oligonucleotides. Such abiotic synthesis is based on a nucleic acid template and a complementary nucleotide or oligonucleotide, which is formed by hydrogen bonding to form a helical complex. The conformation of this complex facilitates the bonding of the 5'-terminal on the substrate molecule to the 3′ or 2′ free hydroxyl groups on the other substrate molecule (Figure 1.5) [53]. The reaction must be carried out at a temperature below the melting point of the helical complex, and the melting point generally increases as the concentration of the reactants and the concentration of ions in the environment increases. This reaction corresponds to the catalysis of DNA or RNA polymerases and DNA ligases in the biological system.

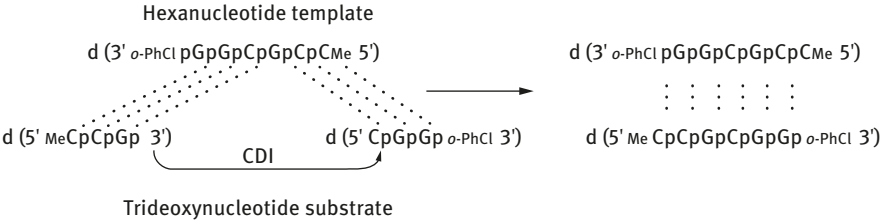


Figure 1.5: Self-catalyzed replications under the guidance of RNA 6-mers templates.

The activated nucleotides, as reactants, may be nucleotides or oligonucleotides. The first example of this reaction is the reaction of the two $(pT)_6$ fragments on the template po1y (dA), resulting in the formation of $(pT)_{12}$ by the water-soluble condensate carbodiimide [54].

Some metal ions have a catalytic effect on template-directed nucleotide condensation. In the presence of Pb^{2+} and Zn^{2+}, po1y (C) can guide the polymerization of

ImpG to produce oligomeric guanosine acid fragments. But in the presence of Pb^{2+}, it mainly forms 2′-,5′-connected non-natural oligomers with chain length up to 10-mers [55]. If 0.01–0.1 mol Zn^{2+} was added, the oligonucleotides linked by 3′-,5′-phosphate diester were mainly produced, and the chain length was up to 30 monomers [56]. If 2-MeImpG was used as the reactant, a 3′-,5′-G oligonomer (chain of up to 50 monomers) was formed with the absence of Zn^{2+} [57]. When po1y (C) reacts with the mixture of 2-MeImpG, 2-MeImpU, 2-MeImpC and 2-MeImpA, the selectivity of po1y (C) to G monomer was 100–500 times that of other monomers [58].

Kiedrowski (1986) [59] utilized a hexanucleotide as a template to bind two deoxynucleotides to each other by base complementary pairing. Under the action of condensation agents, the substrate was condensed to produce a new 6-mers template (Figure 1.5). Another example was the modification of the 3′-hydroxyl group of the dinucleotide substrate to increase the reactivity by using a tetranucleotide molecule as a template to obtain a substrate-condensed tetramer [60] (Figure 1.6).

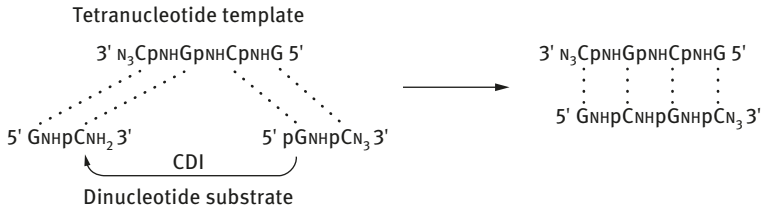


Figure 1.6: Self-catalyzed replications under the guidance of RNA 4-mers templates.

The non-enzyme template-guided synthesis of oligonucleotide follows these four rules:

1. 2′-,5′-oligonucleotides have template activity, but are less active than 3′-,5′-oligonucleotide templates.
2. The reaction requires a high concentration of reactants (10 –25 mmol).
3. The oligomerization of the purine mononucleotides can be carried out on the pyrimidine oligonucleotide template, or conversely, it does not work.
4. Only when the initial pyrimidine nucleotides are 3-mer or longer oligomers, the complementary purine polynucleotide template can guide oligonucleotides.

Besides, there is another pretty important template based on some peptides containing basic amino acids can hydrolyze RNA [61]. These peptides may catalyze the formation of RNA according to the principle of microscopic reversibility. This inference is confirmed by the polymerization of ImpdGpIm catalyzed by po1y (Leu–Lys) in the presence of pH = 6.5, Mn^{2+} and oligonucleotides (18-mers) were obtained [62].

1.4 Irreplaceability of phosphorus in life

As we know that phosphorus is a trace element in the universe, however, why did nature choose phosphorus to be responsible for so many critical roles in biosystem? About this question, Prof. G.M. Blackburn discussed in details [63]. Phosphoric acid has a unique ability among all the elemental oxyacids to form condensed polymers, both chains and rings, which are proven to be phosphorylating agents under mild conditions. They can form directly to phosphate mono- and diester bonds with extreme stability (the half-life of DNA to spontaneous water hydrolysis is 31 million years). It is very important for us that many of them originated at the dawn of life to possess the stability of phosphate esters for genomic purposes with their ready manipulation to develop regulatory and control processes. Besides, the phosphate mono- and diester bonds also have functional susceptibility to rapid manipulation by phosphoryl transfer enzymes. However, other elemental oxyacid esters do not have the required stability (Table 1.5). As Prof. G.M. Blackburn reported that, "since there appears to be no oxyacid from an alternative element in the Periodic Table capable of bridging these two extremes, there is a strong case for the contention that phosphate will be found to command comparable roles in life wherever it exists in the universe" [63].

Table 1.5: Candidate elements and their salts and esters [63].

Element	Oxyacid				Oxyacid esters					
	Oxyacid	pK_a1	pK_a2	pK_a3	Diester Stability /$t_{1/2}$	Diester charge	Bond cleaved	Monoesters tability/ $t_{1/2}$	Monoestr charge	Bond cleaved
Si	H_4SiO_4	9.5	>13	—	< 1 min	0	Si–O	< 1 min	0	Si–O
P	H_3PO_4	2.1	7.2	13.1	10^5 y	−1	P–O	10^{12} y	−2	P–O
V	H_3VO_4	3.2	7.8	12.5	< 1 s	−1	V–O	<< 1 s	−1	V–O
As	H_3AsO_4	2.2	7.0	11.5	< 2 min	−1	As–O	6 min	−2	As–O
S	H_2SO_4	< 0	2.0	—	1.7 h	0	C–O	1,100 y	−1	C–O

1.5 Subject of the origin of life

Although scientists have achieved considerable results in exploring the evolution of life, there is still a long way to go to rationally reveal the essence of life. There are still many problems to be solved in this field. For example, the coupling relationship of nucleic acids and proteins has no reasonable explanations. The origin of genetic code is still unclear. In more specific fields, such as the origin of chirality in the origin of life, protein skeletons only choose L-amino acids (but not β, γ amino acids), and nucleic acid skeletons only choose D-ribose (but not other types of sugars, such as

glucose). It is still no reasonable explanation why DNA and RNA use phosphodiester bond as the basic skeleton and so on. To make a breakthrough in the study of these questions, it is necessary to go through a long time to work, and we need to have more excellent researchers to join this field.

References

[1] Patel B H, Percivalle C, Ritson D J, Duffy C D, Sutherland J D. Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism[J]. *Nature Chemistry*, 2015, *7*, 301–307.

[2] Adcock C T, Hausrath E M, Forster P M. Readily available phosphate from minerals in early aqueous environments on Mars[J]. *Nature Geoscience*, 2013, *6*, 824–807.

[3] Anders E, Grevesse N. Abundances of the elements: Meteoritic and solar[J]. *Geochimica et Cosmochimica Acta,* 1989, *53*, 197–214.

[4] Turner B E, Bally J. Detection of interstellar PN – The first identified phosphorus compound in the interstellar medium[J]. *Astrophysical Journal*, 1987, *321*, L75–L79.

[5] Guelin M, Cernicharo J, Paubert G. Turner B B. Free carbon phosphide (CP) in IRC + 10216[J] *Astronomy and Astrophysics*, 1990, *230*, L9–L11.

[6] Matthews H E, Feldman P A, Bernath P F. Upper limits to interstellar PO[J]. *Astrophysical Journal*, 1986, *312*(1), 358–362.

[7] Schwartz A W. Phosphorus in prebiotic chemistry[M]. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 2006, *361*, 1743–1749.

[8] Zolensky E, Wilson T L, Rietmeijer F J M. Analysis of Interplanetary Dust. In: AIP Conference Proceedings [M]. New York: AIP Press, 1994, 310.

[9] Schwartz A W. The Origins of Life and Evolutionary Biochemistry[M]. Plenum Press, New York, 1974, 435.

[10] Baltscheffsky H, Baltscheffsky M. Early Life on Earth, Nobel Symposium No. 84[M], Columbia University Press, New York, 1994, 81.

[11] Ridgway S T, Wallace L, Smith G R. The 800–1200 Inverse Centimeter Absorption Spectrum of Jupiter. *Astrophys Journal*, 1976, *207*, 1002–1006.

[12] 王文清, Kobayashi K, Ponnamperuma C. 甲烷、氮、三氢化磷与水混合物的预生物合成.[J] 北京大学学报（自然科学版), 1984, *6*, 36–44.

[13] 王文清. 三氢化磷、甲烷、氮与水混合物的前生物合成一PH_3在化学进化中的作用.[J] 科学通报, 1984, *29*, 1344–1344.

[14] WANG W Q（王文清）. Prebiotic Synthesis In a Mixture of Phosphine, Methane, Nitrogen and Water [J]. *Science Bulletin* (Kexue Tongbao). 1985, 30(2), 281.

[15] Fox S W, Harada K. Synthesis of uracil under conditions of a thermal model of prebiological chemistry [J]. *Science*, 1961, *133*, 1923–1924.

[16] Martin F, David N, Jiří Š. High-energy chemistry of formamide: a unified mechanism of nucleobase formation [J]. *PNAS*, 2015, 112 (3): 657–662.

[17] Ponnamperuma C, Mariner R, Sagan C. Formation of adenosine by ultraviolet irradiation of a solution of adenine and ribose [J]. *Nature*, 1963, *198*, 1199–1200.

[18] Ponnamperuma C, Kirk P. Synthesis of deoxyadenosine under simulated primitive earth conditions [J]. *Nature*, 1964, *203*, 400–401.

[19] Schramm G. Synthesis of nucleosides and polynucleotides with metaphosphte esters. In: Origins Prebiol. Systems Their Mol. Matrices, Proceedings of Conferrence, Wakulla Springs, FL, 1963, 299–309.

[20] Carbon J A. Schramm deoxyadenosine synthesis [J]. *Chemical Industry*, 1963, 529–530.
[21] Ponnamperuma C, Mack R. Nucleotide synthesis under possible primitive earth conditions [J]. *Science*, 1965, *148*(3674), 1221–1223.
[22] Beck A, Lohrmann R, Orgel L E. Phosphorylation with inorganic phosphates at moderate temperatures [J]. *Science*, 1967, *157*, 952–952.
[23] Lohrmann R, Orgel L E. Prebiotic synthesis: Phosphorylation in aqueous solution [J]. *Science*, 1968, *161*, 64–66.
[24] Schwartz A, Ponnamperuma C. Phosphorylation of adenosine with linear polyphosphate salts in aqueous solution [J]. *Nature*, 1968, *218*, 443–443.
[25] Waehneldt T V, Fox S W. Phosphorylation of nucleosides with polyphosphoric acid [J]. *Biochimica et Biophysica Acta*, 1967, *134*(1), 1–8.
[26] Miller S L, Parris M. Synthesis of pyrophosphate under primitive earth conditions [J]. *Nature*, 1964, *204*(4965), 1248–1250.
[27] Saffhill R. Selective phosphorylation of the cis-2′,3′-Diol of unprotected ribonucleosides with trimetaphosphate in aqueous solution [J]. *Journal of Organic Chemisty*, 1970, *35*(9), 2881–2883.
[28] Yamagata Y, Watanabe H, Saitoh M et al. Volcanic production of polyphosphates and its relevance to prebiotic evolution [J]. *Nature*, 1991, *352*, 516–519.
[29] Ferris J P, Goldstein G, Beaulieu D J. Chem. Evolution. IV. An evaluation of cyanovinyl phosphate as a prebiotic phosphorylating agent [J]. *Journal of the American Chemical Society*, 1970, *92*(22), 6598–6603.
[30] Moravek J Formation of oligonucleotides during heating of a mixture of uridine 2′(3′)-phosphate and uridine [J]. *Tetrahedron Letters*, 1967, (18), 1707–1710.
[31] Tsuhako M, Fujimoto M, Ohashi S. Phosphorylation of inosine with cyclo-triphosphate [J]. *Chemistry Letters*, 1981, 155(7), 849–852.
[32] Tsuhako M, Fujimoto M, Ohashi S. Phosphorylation of nucleosides with sodium cyclo-triphosphate [J]. *Bulletin of the Chemical Society of Japan*, 1984, *57*(11), 3274–3280.
[33] Feldmann W. Das Trimetaphosphat als Triphosphorylierungsmittel für Alkohole und Kohlenhydrate in Wäßriger Lösung. Seine Sonderstellung unter den kondensierten Phosphaten [J]. *Chemische Berichte*, 1967, *100*(12), 3850–3860.
[34] Tsuhako M, Sueyoshi C, Baba Y et al. Phosphorylation of glucose with cyclo-triphosphate [J]. *Chemical Letters*, 1987, 187(7), 1431–1434.
[35] Yamanaka J, Inomata K, Yamagata Y. Condensation of oligoglycines with trimeta- and tetrametaphosphate in aqueous solutions [J]. *The Journal Origins of Life and Evolution of Biospheres*, 1988, *18*(3), 165–178.
[36] Fox S W. The origins of prebiological systems and of their molecular matrices [M]. New York, Academic Press, 1965, 289.
[37] Feldmann W. Ringaufspaltung des Trimetaphosphations, $P_3O_9^{3-}$, durch Aminosäuren in Wäßriger Lösung; Über die Knüpfung von Peptid-bindungen mit $P_3O_9^{3-}$ [J]. *Zeitschrift für Physikalische Chemie*, 1969, *9*(4), 154–155.
[38] Rabinowitz J. Preparation and transformation of Esters. LXXXIII. Condensation and (or) phosphorylation reactions in aqueous solution of organic compounds containing hydroxy, carboxy, amino and other functional groups with linear or cyclic phosphates [J]. *Helvetica Chimica Acta*, 1969, 52(8), 2663–2671.
[39] Rabinowitz J, Flores J, Krebsbach R et al. Peptide formation in the presence of linear or cyclic polyphosphate [J]. *Nature*, 1969, 224, 795–796.
[40] Hulshof J, Ponnamperuma C. Prebiotic condensation reactions in an aqueous medium: A review of condensation agents [J]. *Origins of Life*, 1976, *7*(3), 197–224.
[41] Schramm G, Grötsch H, Pollmann W. Non-enzymatic synthesis of polysaccharides, nucleosides, and nucleic acids [J]. *Angewandte Chemie*, 1961, *73*, 619–619.

[42] Schwartz A W, Fox S W. Condensation of cytidylic acid in the presence of poly-phosphoric acid [J]. *Biochimica et Biophysica Acta*, 1967, *134*(1), 9–16.
[43] Contreras G, Espejo R, Mery E et al. Polymerization of ribonucleotides by γ-radiation [J]. *Biochimica et Biophysica Acta*, 1962, *61*, 718–727.
[44] Steinman G, Capolupo A A. Chemical plausibility of dicyanamide as a primordial reactant [J]. *Currents in Modern Biology*, 1969, *2*(6), 295–297.
[45] Ibanez J D, Kimball A P, Oró J. Possible prebiotic condensation of mononucleotides by cyanamide [J]. *Science*, 1971, *173*, 444–446.
[46] Odom D G, Brady J T, Oró J. The prebiotic synthesis of deoxythymidine oligonucleotides. III. Acid salt reactions [J]. *Journal of Molecular Evolution*, 1976, *7*(2), 151–157.
[47] Lohrmann R. Formation of nucleoside 5′-phosphoramidate under potentially prebiological conditions [J]. *Journal of Molecular Evolution*, 1977, *10*(2), 137–154.
[48] Cairns-Smith A G. The first organism [J]. *Scientific American*, 1985, *252*(6), 74–82.
[49] Ding P Z, Kawamura K, Ferris J P. Oligomerization of uridine phosphorimidazolides on montmorillonite: A model for the prebiotic synthesis of RNA [J]. *Origins of Life and Evolution of Biospheres*, 1996, *26*(2), 151–171.
[50] Ferris J P, Ertem G. Oligomerization of ribonucleotides on montmorillonite: Reaction of the 5′-phosphorimidazolide of adenosine [J]. *Science*, 1992, 257, 1387–1389.
[51] Ferris J P, Ertem G. Oligomerization of ribonucleotides: The reaction of 5′-phosphorimidazolide of nucleosides on montmorillonite and other minerals [J]. *Origins of Life*, 1992, 22(6), 369–381.
[52] Lohrmann R, Orgel L E. Formation of P_1, P_2 -dinucleoside 5′-pyrophosphates under potentially prebiological conditions [J]. *Journal of Molecular Evolution*, 1978, *11*(1), 17–23.
[53] Orgel L E. Mini review RNA catalysis and the origins of life [J]. *Journal of Theoretical Biology*, 1986, 123, 127–149.
[54] Naylor R, Gilham P T. Studies on some interactions and reactions of oligonucleotides in aqueous solution [J]. *Biochemistry*, 1966, *5*(8), 2722–2728.
[55] Lohrmann R, Orgel L E. Efficient catalysis of polycytidylic acid-directed oligoguanylate formation by lead [J]. *Journal of Molecular Biology*, 1980, *142*(4), 555–567.
[56] Bridson P K, Orgel L E. Catalysis of accurate poly(C)-directed synthesis of 3′- 5′-linked oligoguanylates by zinc [J]. *Journal of Molecular Biology*, 1980, *144*(4), 567–577.
[57] Inoue T, Orgel L E. Substituent control of the poly(C)-directed oligomerization of guanosine 5′-phosphoroimidazolide [J]. *Journal of the American Chemical Society*, 1981, *103*(25), 7666–7667.
[58] Inoue T, Orgel L E. Oligomerization of (Guanosine 5′-phosphor)-2-methylimidazolide on Poly(C), An RNA polymerase model [J]. *Journal of Molecular Biology*, 1982, *162*(1), 201–217.
[59] Von Kiedrowski G. A self-replicating hexadeoxynucleotide [J]. *Angewandte Chemie International Edition*, 1986, *25*(10), 932–935.
[60] Zielinski W S, Orgel L E. Oligomerization of activated derivatives of 3′-Amino-3′-deoxyguanosine on poly(C) and poly(dC) templates [J]. *Nucleic Acids Research*, 1985, *13*(7), 2469–2486.
[61] Barbier B, Brack A. Conformation-controlled hydrolysis of polynucleotides by sequential basic polypeptides [J]. *Journal of the American Chemical Society*, 1992, *114*(9), 3511–3515.
[62] Barbier B, Visscher J, Schwartz A W. Polypeptide-assisted oligomerization of nucleotide analogs in dilute aqueous-Solution [J]. *Journal of Molecular Evolution*, 1993, *37*(6), 554–558.
[63] Bowler M W, Cliff M J, Waltho J P, Blackburn G M. Why did nature select phosphate for its dominant roles in biology [J]? *New Journal of Chemistry*, 2010, *34*, 784–794.

2 Why nature chose α-amino acids?

It was found that only *N*-phosphoryl α-amino acids, not the β- or γ-amino acids, could be activated by themselves to form biological compounds. Therefore, it could be said that the phosphorus chose α-amino acids.

2.1 Amino acids and the origins of life

Protein is the basic material of life and plays key roles in all biological processes, which is a major contributor to biological functions and life activities. The basic building block of proteins is the amino acid, which forms the linear peptide chain through the chemical amide bonds as shown in Figure 2.1. Twenty-two amino acids are naturally incorporated into polypeptides and are called proteinogenic amino acids. After billions of years of evolution, the life system has formed a very complete molecular regulation mechanism. The peptide chain skeleton of the protein is strictly synthesized according to the DNA information of the genetic code with three bases that determine an amino acid. How did the genetic code origin? Why all of the protein-derived amino acids are L-configurations with the α-amino groups? The answer to these basic scientific questions of human will provide important clues to the origin of life [1].

A hundred years ago, Charles Darwin published his very famous book *Origin of Species*. A great many of vivid experimental evidence has been used to show how life can go from simplicity to complexity in order to determine the origin of species. Rationally, the origin of species should first be based on the origin of life. So far, the origin of life is one of the top 10 unsolved mysteries of chemistry on Earth. Darwin described the origin of life in a single sentence at the end of this great book: "Probably all the organic beings which have ever lived on this Earth have descended from some one primordial form, into which life was first breathed" [2]. Meanwhile, Darwin proposed that "In some warm little pond, with all sorts of ammonia and phosphoric salts, light, heat, a protein compound was chemically formed." He postulated that phosphorus and nitrogen elements play a key role in the origin of life. Indeed, phosphorus plays an important role in today's life, which has been evolved after hundreds of millions or even billions of years of constant selection, evolution and accumulation. As noted by biochemist F.H. Westheimer, "Why nature chose phosphates," some intrinsic properties of phosphorus were determined as one of the central biological elements [3]. However, whether phosphorus is already playing a role in the birth of life on Earth or has been chosen in the evolution process of life? Did the special chemical properties of phosphorus promote or accelerate the birth of life on Earth? How to form the active phosphate such as cyclic polyphosphate and polyphosphate? Did phosphorus element also play a decisive role in assembling small molecules into functional

https://doi.org/10.1515/9783110562552-002

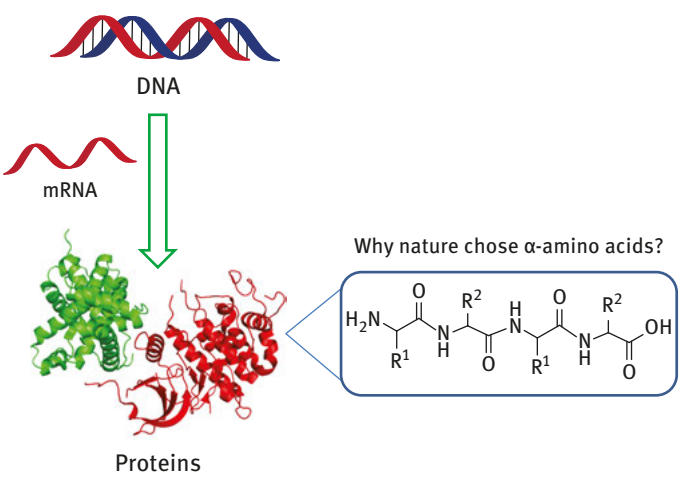


Figure 2.1: α-Amino acids are the basic building blocks for biosynthesis of natural proteins.

biomolecules? There are so many problems to which we cannot find the answers, which date back to billions of years ago. The answers to these questions can only depend on the analysis and summary of today's phenomena of life. To construct various models, especially the chemical small molecule model, reasonable speculation on the origin of life under the former prebiotic conditions is made in order to further reveal the nature of the origin of life. A growing body of research suggests that the phosphorus-chemical process in the life system may have begun to play a unique and important role before the beginning of life. As professor De Duve C said, "Life is basically organized around phosphorus" [4]. This chapter will focus on the dynamic formation of peptides activated by phosphorus reagents with α-, β- and γ-amino acids under prebiotic conditions, which may provide some scientific basis for the natural selection of α-amino acids.

2.2 General structures of amino acids and the biological functions

More than 500 amino acids have been found so far, but the protein skeleton strictly consists of α-amino acids. Amino acids can be classified according to the different physical and chemical properties, such as polarity, pH and the chemical structure of side chains. For example, amino acids can be divided into acidic, alkaline and neutral amino acids, or be divided into α- (alpha), β- (beta), γ- (gamma) and δ- (delta) amino acids based on the position of amino groups. β-Amino acid still plays an important role in today's life activities, such as β-alanine, which is a naturally occurring amino acid. β-Alanine is produced by hydrolysis of dihydrogen uracil and muscle peptide

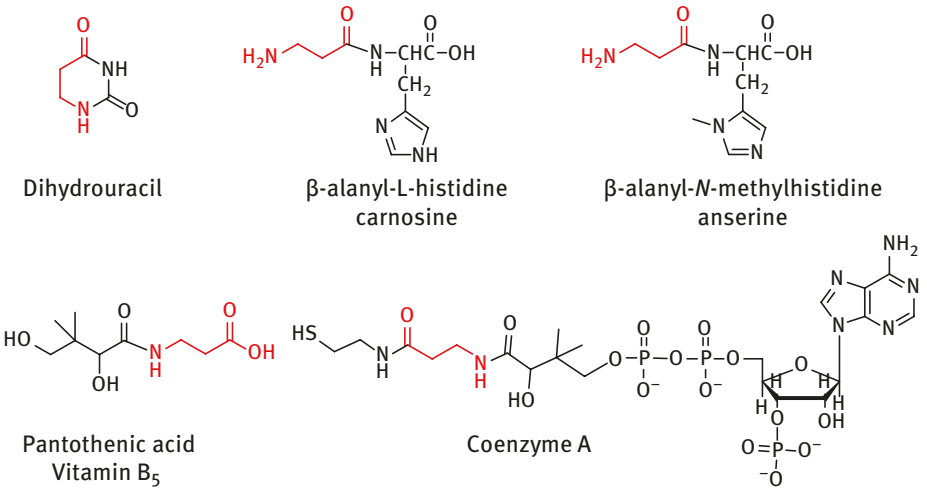


Figure 2.2: Metabolites containing the β-alanine residue.

in the body, as shown in Figure 2.2. Several metabolites contain the core structure of β-alanine. Carnosine (β-alanyl-L-histidine) is a β-alanine and histidine dipeptide; with high content in muscle and brain tissue, some biological functions have been revealed, and the preliminary study found that Carnosine and anserine can remove the reactive oxygen species as antioxidant and anti-aging reagent, which are widely used in cosmetics and health food. Furthermore, Carnosine could also be used as a potential drug candidate for a variety of degenerative disease, such as diabetes, chronic kidney disease and the Alzheimer's disease. Another metabolite that contains β-alanine is the pantothenic acid, also known as vitamin B5, which is a key component for the synthesis of coenzyme A molecule. In addition, γ-aminobutyric acid (GABA) is an important neurotransmitter widely distributed in animal central nervous tissue.

However, even if the β-alanine and γ-aminobutyric acid as endogenous metabolites play important biological functions in the human body, only α-amino acids could be incorporated into proteins. Before the origin of life, what mysterious force chose α-amino acid as raw protein structural unit for polymerization to produce biological macromolecules without enzymatic catalysis in the process of evolution? The dynamic process of peptide chain formations maybe one of the best model for determination of the possible answers for abovementioned questions.

2.3 Prebiotic formation of amino acids

Although hundreds of amino acids have been found in biological systems, all the natural amino acids derived from proteins have α-amino and carboxyl groups.

These are the only differences between the chemical structure of side chains, which have different physical and chemical properties, such as solubility, acid, alkaline, optical properties and so on, thus endowed with different structures and functions of the protein. Amino acids could be prebiotically synthesized through a variety of natural conditions, such as the simulation of lightning and the natural extreme conditions such as volcanic eruptions. In addition, a variety of amino acids have been determined in moon rocks and Martian meteorites by using modern analytical instruments, such as high performance chromatography coupled with mass spectrometry technology. Based on such abundant experimental evidence, amino acids as the most basic substances of life are no longer constraints to the origin of life.

In 1953, scientist Miller in the United States had carried on the famous "spark discharge" experiment, using a reducing gas mixture, such as ammonia, hydrogen, methane and water, under the prebiotic conditions, and successfully detected biological substances, including 10 kinds of proteinogenic amino acids such as glycine, serine, alanine and valine [6]. The successfully prebiotic synthesis of amino acids provide direct scientific basis of chemical origin theory. The research of the origin of life since then has entered a new era. The presence of PH_3 is then detected in the atmosphere of Jupiter and Saturn. The findings suggest that PH_3 may exist in the original atmosphere and may be involved in synthesizing biological organic small molecules. Under the same condition, Ponnamperuma C and his collaborators used the original simulated atmospheric CH_4+NH_3+H_2+H_2O containing PH_3 and the same gas mixture with no PH_3 [7]. The results showed that in the absence of PH_3, only 6 kinds of simple amino acids have been detected. In contrast, 19 kinds of amino acids, which contain more complex amino acids, such as phenylalanine, proline and serine, have been successfully identified with the catalysis of PH_3. This experiment strongly suggests that phosphorus may play an important role in the origin of amino acids. In 2008, a student of SJ Miller, Baba reanalyzed Miller's experimental samples, using modern analysis methods, such as chromatography with high resolution mass spectrometry technology, 22 kinds of amino acids and organic amine compounds are found[8]. In particular, besides α-amino acids, the experiment also produced a variety of beta and gamma amino acid, for example, β-alanine, β-aminobutyric acid and γ-aminobutyric acid. In addition, space meteorites are widely believed to be an important source of the Earth's amino acids. An analysis of the composition of various meteorites found that there was a large amount of organic matter in meteorites, including a variety of amino acids. More than 40 years of research, more than 80 kinds of amino acids, including β-alanine, β-aminobutyric acid and γ-aminobutyric acid and even ε-amino-*n*-caproic acid (EACA), have been found in a piece of meteorites that fell in Australia in 1969 by using different approaches (Figure 2.3). In addition, ocean hot springs are another possible pathway for amino acids.

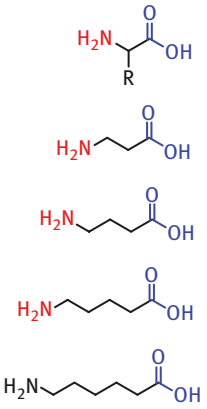


Figure 2.3: Chemical structures of amino acids.

2.4 Prebiotic formation of peptides

The life system has evolved to perform protein synthesis using complex biological machines with highly specific enzymes. The key molecular chemical structure of protein biosynthesis, α-amino acids-AMP, is phosphoric acid anhydride. However, how did the first bioactive polypeptide or protein come from? The synthesis of polypeptide or protein in the prebiotic condition requires a completely different approach. Under the prebiotic conditions without high catalytically active enzymes, amino acids can only be obtained by physical or chemical process such as amino acid dehydration forming amide bond, thus forming a complex structure of peptides that has the function of the protein to lay the material foundation for the final product. So why did nature choose α-amino acids as basic units in protein synthesis in the life system today? The answer to this question is that amino acids with different chemical structures could be activated by the specific enzyme in the dynamic process of condensation in the formation of peptide bonds. In the prebiotic condition, the amino acid undergoes no enzymatic catalysis, which can only activate the amino acid molecules through chemical energy and form amide bonds.

There are many possible pathways for the formation of peptide bonds under the prebiotic conditions, which are mainly divided into thermodynamic processes and kinetic control processes. The change of thermodynamic processes such as inorganic salt, high temperature, pH and wet reaction conditions can promote the condensation of amino acids into peptides. However, the efficiency of peptide formation is low because of the limited concentration of amino acids. Thus, the dynamic process of amino acid under the catalysis of activation reagent into peptide may play a more important role. Many possible prebiotic activation reagents have been subsequently found [9, 10], such as volcanic eruptions of cosine gas, which can efficiently activate amino acids in the production of amino acids – N-carboxyanhydride (NCA) active intermediates, which can

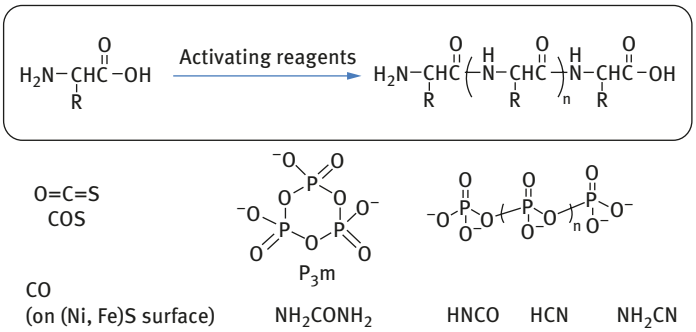


Figure 2.4: Chemical structures of prebiotic reagents for peptide formation.

be subsequently polymerized to generate peptides. Volcanic activity can also produce polyphosphate and cyclic polyphosphate, especially trimetaphosphate (P_3m), which can act as an efficient catalyst in alkaline aqueous solutions of peptides and amino acids, such as the dipeptide yield of glycine can be up to 30%. In addition, other activated reagents such as CO, urea, HCN and NH_2CN were also proposed by scientists (Figure 2.4).

2.5 Phosphorus chemistry for prebiotic peptide formation

Proteins are produced by using α-amino acid as the basic building blocks for amide bond formation, which require participation of multiple enzymes and finally complete the formation of a polypeptide chain in ribosomes. The mechanism of protein biosynthesis is formed by hundreds of millions of years of evolution and selection. The vital origin of macromolecules with biological functions from the small molecules has been one of the key topics in the research of life origins. By simulating a variety of possible prebiotic conditions, amino acids can be condensed to form a variety of polypeptide chains with different lengths, suggesting multiple possibilities for the formation of primitive polypeptides and proteins. Phosphorus is ubiquitous in today's life system and is the center of life activity. With the catalysis of inorganic phosphate and organic phosphorus reagents, condensation of amino acids can be very effective to promote the formation of peptide bonds under prebiotic conditions. However, due to the extremely low concentration of phosphorus that can be easily formed by insoluble inorganic salts with metal ions in aqueous solution, limits the role of phosphorus to some extent in the process of the origin of life. In recent years, with the continuous emergence of new research, more and more studies have proved the richness and diversity of the content of water-soluble active phosphorus elements. Natural conditions such as volcanic activities, meteorites and natural lightning can produce a large number of soluble phosphates [11, 12]. As shown in Figure 2.5, the main form of water-soluble or reactive inorganic phosphate is found in nature.

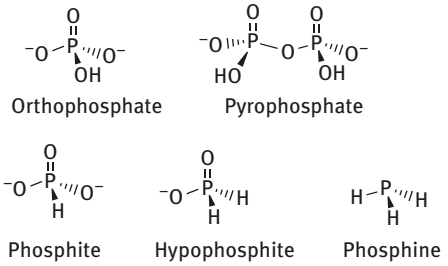


Figure 2.5: The chemical structures of inorganic phosphates in nature.

2.5.1 Reaction of amino acids with trimetaphosphate

Volcanic activity is considered to be a common phenomenon in prebiotic conditions, even though volcanoes and submarine hot springs are still widely distributed on Earth today. It is interesting to note that in 1991, Yamagata Y discovered the existence of polyphosphate in samples produced by volcanic activity, which also includes a small amount of partial P_3m. Furthermore, a mixture of a variety of phosphate could also be obtained, such as pyrophosphate, tripolyphosphate and partial tripolyphosphate under the experimental simulation of volcanic conditions, suggesting the source of the polyphosphates [13]. Pyrophosphate, P_3m and a small amount of polyphosphate can also be generated by the polymerization of free radical reaction. It is estimated that the annual output of polyphosphate in neutral conditions can exceed 10^9 mol/year [14]. In summary, soluble polyphosphate can not only be obtained from extraterrestrial meteorites, but also through natural activities in the Earth itself. As Prof. Schwartz AW said: "Phosphorus problem is no longer the stumbling block which it was once thought to be," phosphorus is no longer a limiting factor in prebiotic chemistry. Inorganic phosphates, such as cyclic polyphosphates and polyphosphates, may play an important role in the origin of life. In the induction of inorganic phosphate, amino acids and nucleosides can be polymerized into bioactive macromolecules, such as polypeptides, proteins or nucleic acids under the prebiotic conditions. In this critical process, inorganic phosphate can effectively contract and catalyze the production of biomolecules. For example, J Rabinowitz found that P_3m can be an effective catalyst for peptide formation, especially the highest efficiency has been obtained with glycine and alanine under the conditions of weak base, low temperature and low concentration in an aqueous solution [15].

As shown in Figure 2.6, the possible pathways for the reaction of P_3m with amino acids and organic small molecules have been proposed and a five-membered cyclic intermediate was formed by the reaction of amino acids with P_3m. P_3m catalyzes various amino acids to form a diverse polypeptide library. However, α-, β- and γ-amino acids exhibit distinctly reactive properties during this process [16]. By liquid chromatography-mass spectrometry (LC-MS) analysis, it was found that the mixture

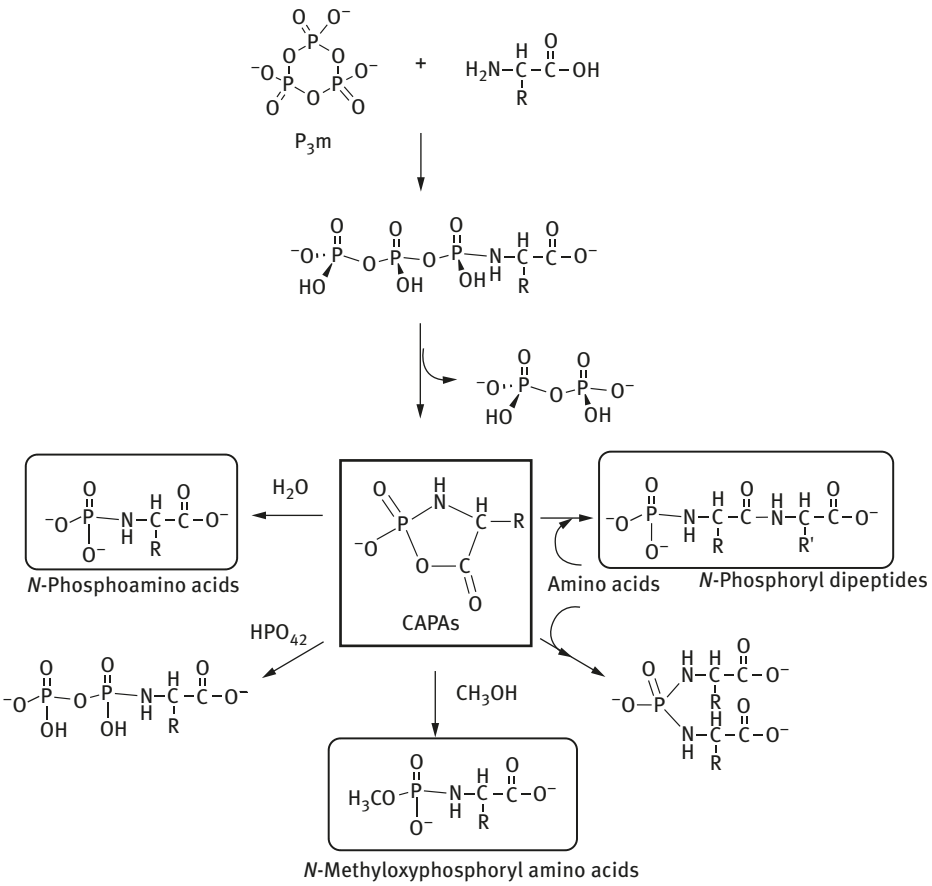


Figure 2.6: The prebiotic reactions of trimetaphosphate treated with amino acids.

of glycine and alanine treated with P_3m generated a total of four kinds of dipeptides, as shown in Figure 2.7, such as Gly–Gly, Gly–Ala, Ala–Gly and Ala–Ala with different yields that have been quantified through integration of the liquid chromatography peaks, suggesting that the efficiency of peptide formation of specific α-amino acid is also different.

However, only two dipeptides with glycine as the *N*-terminal residue, including Gly–Gly and Gly–β–Ala, were produced with different yields as shown in Figure 2.8 further suggesting that β-Alanine cannot be activated by P_3m in the amide bond formation.

Figure 2.9 demonstrated the reaction mixtures of glycine and γ-amino butyric acid treated with P_3m by using LC-MS. The reaction mixture also contains only two kinds of dipeptides, Gly–Gly and Gly–γ–ABA dipeptides, in which only glycine is present in the *N*-terminal of the dipeptide with similar condensation efficiency. A possible mechanism for the reaction of P_3m and amino acids was proposed as shown

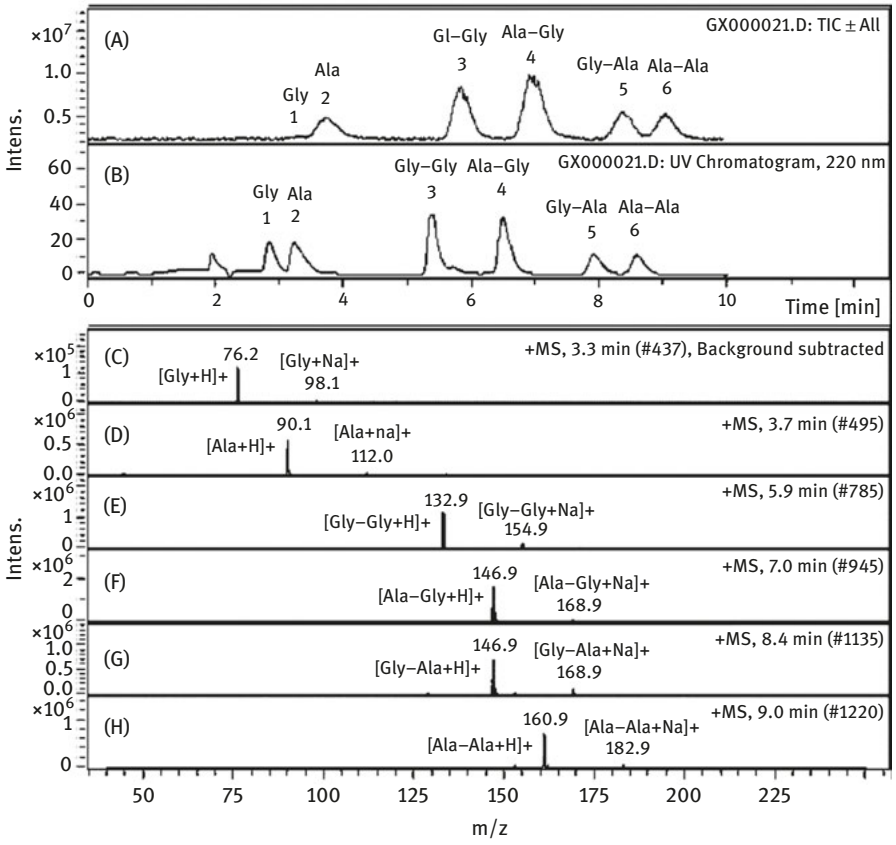


Figure 2.7: Analysis of reaction products of glycine and L-alanine treated with P_3m by using LC-MS: (a) Total ion current (TIC) chromatogram, (b) ultraviolet chromatogram at 220 nm, (c) peak 1, glycine, (d) peak 2, L-alanine, (e) peak 3, Gly–Gly, (f) peak 4, Ala–Gly, (g) peak 5, Gly–Ala and (h) peak 6, Ala–Ala.

in Figure 2.10. First, amino group of amino acid attacks the phosphorus atoms of P_3m, forming triphosphorylated amino acids. Then, the carboxylic acid of the amino acid attacked the α-phosphorus atom to form the pentacoordinated five-member cyclic intermediate with a molecule of pyrophosphate group leaves. The another molecule of amino acid then attacks the carbonyl group of cyclic intermediates, which hydrolyzes to produce dipeptide and phosphate. The five-membered cyclic mixed anhydride can be easily produced by α-amino acids, but it is difficult to form six or seven membered cyclic mixed anhydride intermediates for β- and γ-amino acid respectively. For mixture of amino acids, only with α-amino acids, such as glycine and alanine, can be activated to produce cross small peptide. In conclusion, the inorganic phosphates, such as P_3m, effectively select α-amino acid as the basic building blocks for the formation of polypeptide, rather than β- and γ-amino acids with similar chemical structures, suggesting that important roles have been performed for the origin of life.

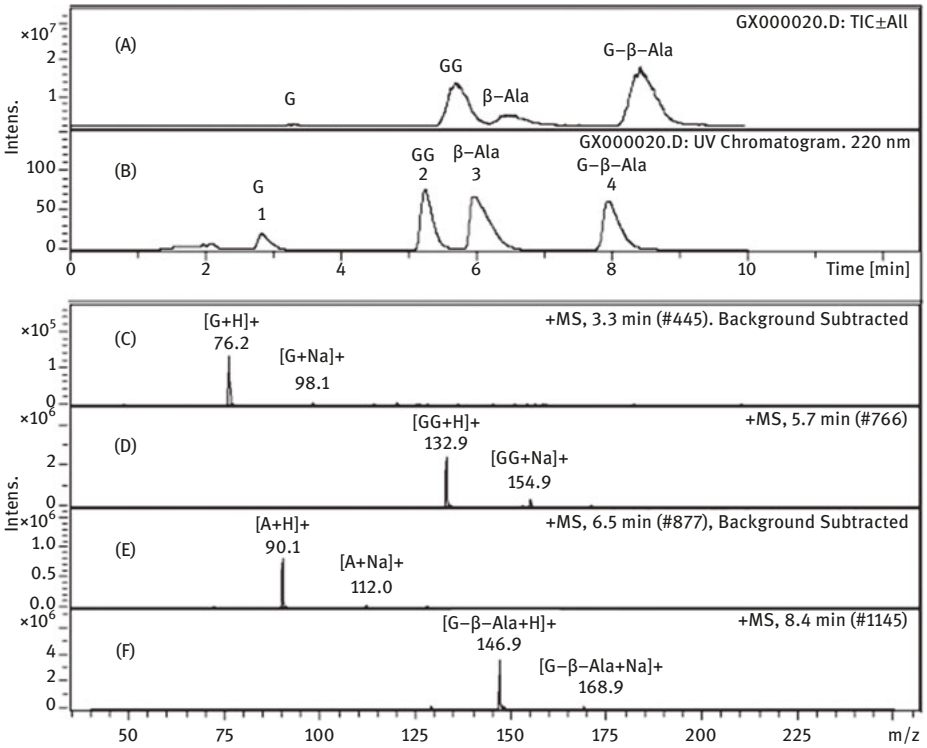


Figure 2.8: Analysis of reaction products of glycine and β-alanine treated with P_3m by using LC-MS: (a) Total ion current (TIC) chromatogram, (b) ultraviolet chromatogram at 220 nm, (c) peak 1, glycine, (d) peak 2, Gly–Gly, (e) peak 3, β-alanine and (f) peak 4, Gly–β–Ala.

2.5.2 Peptide formation of *N*-phosphoryl amino acids

N-Phosphoryl amino acid as the ancient small molecular fossil can be easily produced under prebiotic conditions. *N*-Phosphoryl amino acid can be classified into three categories with different substitutions of phosphoryl groups: (1) *N*-dialkyoxylphosphoryl amino acids (*N*-DAPAAs); (2) *N*-monoalkoxylphosphoryl amino acids (*N*-MAPAAs) [17]; (3) *N*-phospho amino acid (*N*-PAAs). The chemical properties of *N*-DAPAAs have been systematically investigated and several interesting reactions have also been determined [17]. For example, α-amino acid can be activated by *N*-terminal phosphorylation and self-assembled to form oligopeptides. The five-membered cyclic pentacoordinate phosphate intermediates of *N*-phosphorylated amino acids were successfully captured by using ^{31}P NMR in anhydrous organic solvents. In addition, it was found that *N*-phosphoryl β- and γ-amino acids were fairly stable under the same conditions, and no peptides have been produced. There is no new ^{31}P NMR peak formed for *N*-phosphoryl β-alanine or *N*-phosphoryl γ-aminobutyric acid in pyridine with 79 hours incubation as shown in Figure 2.11. However, the situation is very different

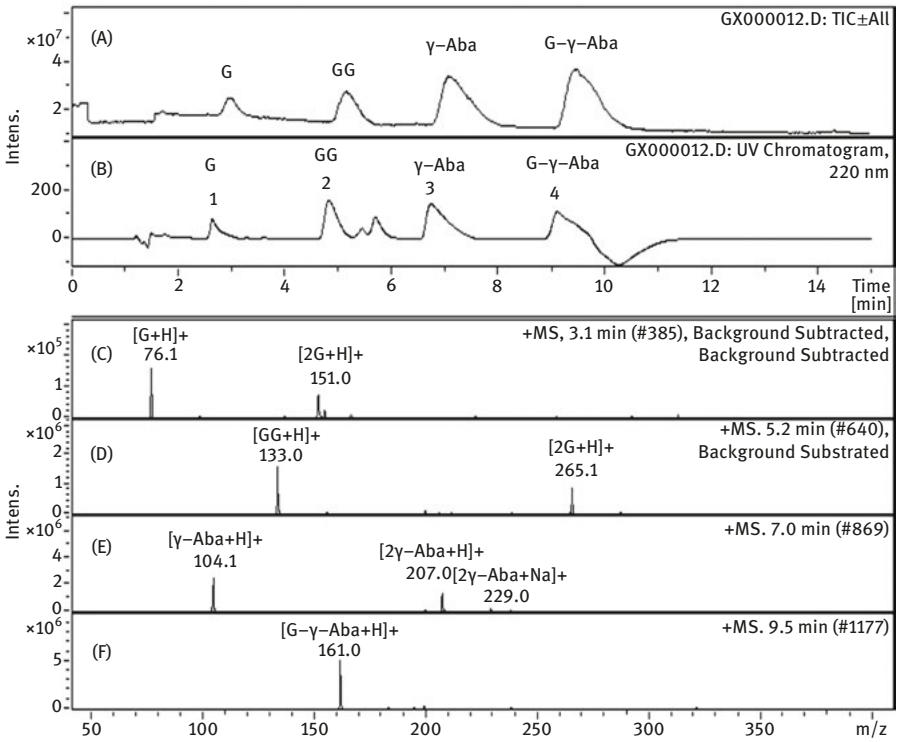


Figure 2.9: Analysis of reaction products of glycine and γ-amino butyric acid treated with P_3m by using LC-MS: (a) Total ion current (TIC) chromatogram, (b) ultraviolet chromatogram at 220 nm, (c) peak 1, glycine, (d) peak 2, Gly–Gly, (e) peak 3, γ-Aba and (f) peak 4, Gly–γ–Aba.

for DIPP-α-Ala. It was found that *N*-DIPP-α-Ala was very active because several new ^{31}P NMR peaks with high intensity are produced at the same concentration in the solvent of water, pyridine and chloroform.

The abovementioned results indicate that *N*-phosphorylation is highly selective in the prebiotic synthesis of peptides with α-, β- and γ-amino acids. It is likely that the phosphorus chooses α-amino acid. Furthermore, it was found that *N*-DAPPs can not only self-assemble into peptides, but also simultaneously form nucleotides and nucleic acids under mild conditions. A new theory of coevolution of proteins and nucleic acids has been proposed as "*N* – phosphoryl amino acid could be considered as the chemical co-evolution model for the origins of nucleic acids and proteins in the process of the origin of life." In addition, the N–O transfer reaction of phosphoryl group in *N*-diisopropylphosphoryl serine is also studied as a small chemical model for the study of phosphoryl group migration catalyzed by protein kinases. The abovementioned small model, containing phosphoryl group, carboxyl group and amino acid side chains, can be investigated as an integrative system for exploring the origins of enzymatic functions, thus *N*-phosphorus amino acid as the smallest structural unit of phosphorus acylation protein can also be called "mini-enzyme" [18].

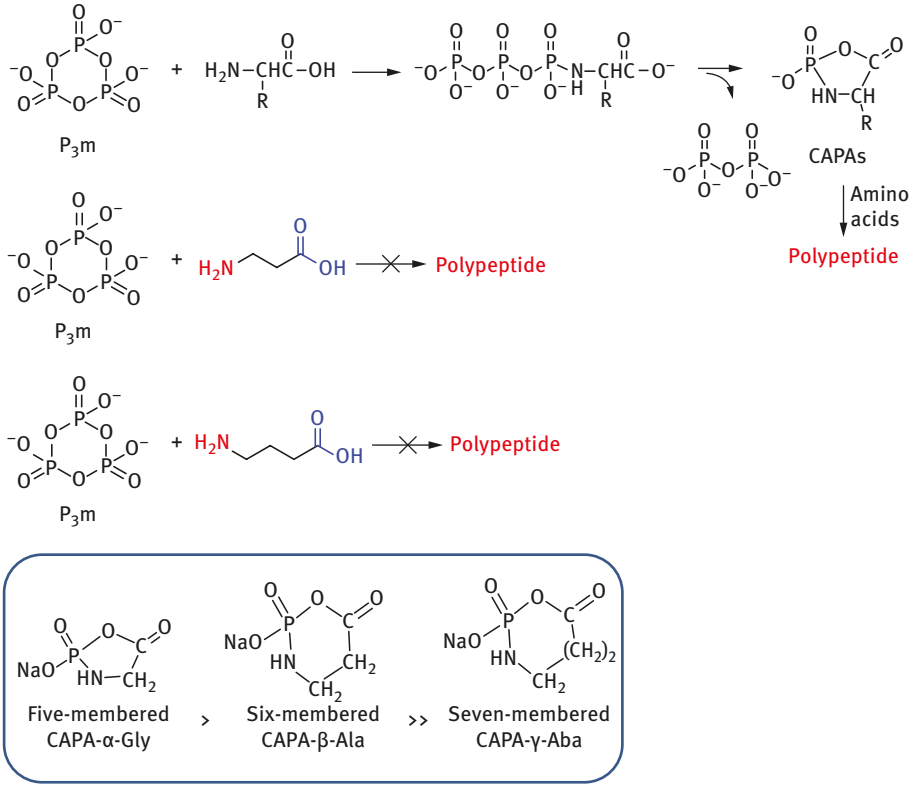


Figure 2.10: The possible chemical mechanism of α-, β- and γ-amino acids treated with P_3m.

Both of the C-terminal carboxyl group and N-terminal amino group of amino acids can be phosphorylated, as shown in Figure 2.12. The activation of amino acid in organisms often adopts C-terminal phosphorylation to form mixed anhydrides under the catalysis of specific enzymes with energy from hydrolysis of high-energy pyrophosphate bond of ATP. However, amino group of amino acid in aqueous solution is very difficult to react with ATP without enzyme catalysis. In the alkaline aqueous solution, N-phosphorylated amino acids can be easily formed with the attack of amino group with inorganic polyphosphates. The chemistry of adenosine phosphate–carboxylic acid anhydride of amino acid has been extensively investigated and many compounds containing phosphate–carboxylic acid mixed anhydride keys were synthesized as study models due to their importance in the biosynthesis of proteins. It is interesting to note that the *N*-monoalkyloxylphosphoryl amino acids are structurally isomers of aminoacyl adenosine, which are only phosphoryl groups at opposite ends of the amino acids. What are the differences in the chemical nature of these two different phosphorylation reagents? The answer may be enlightening to the energy metabolism process in the life system, molecular mechanism of protein kinase catalyzing phosphorylation migration and molecular catalysis.

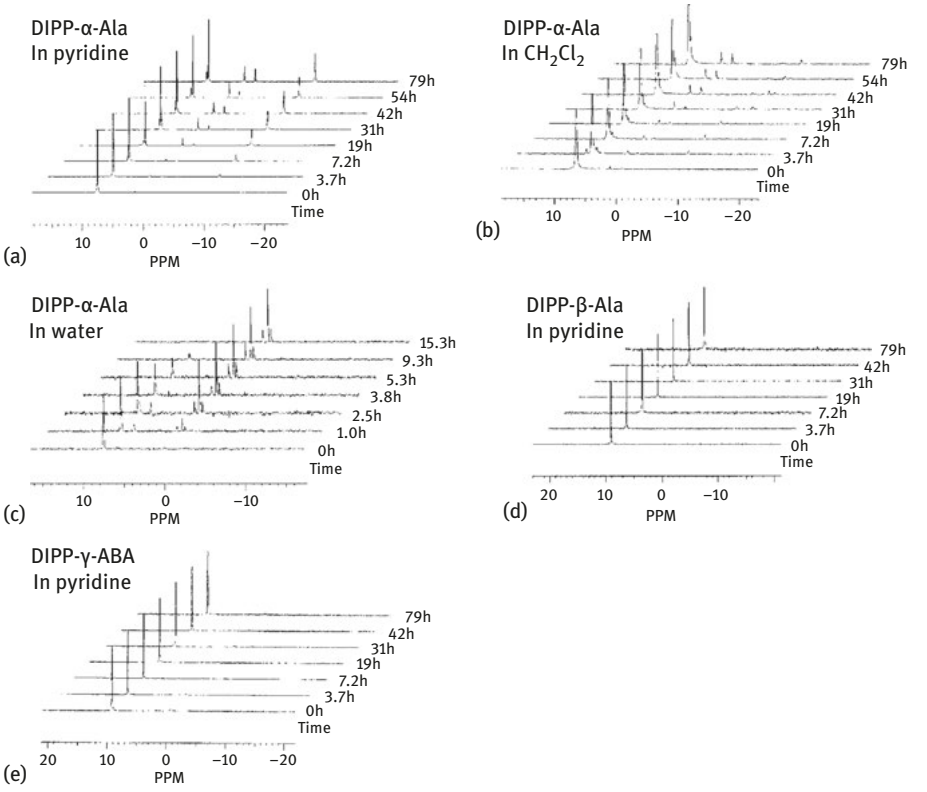


Figure 2.11: The reaction of *N*-phosphorylated amino acids in different solvents by using ^{31}P NMR.

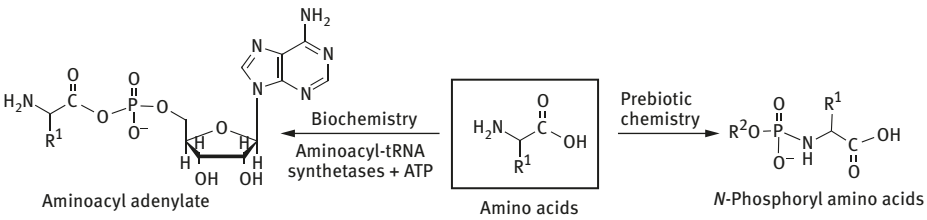


Figure 2.12: The biological and prebiotic activation of amino acids at C- or N-terminus.

2.5.3 Peptide formation mediated by organic phosphorus reagents

Organophosphorus reagents can also be used as a very effective strategy for construction of polypeptides in organic synthesis. Many organophosphorus reagents have been successfully applied to form amine bonds in organic solvents, such as phosphorus oxychloride, phosphorus trichloride and benzene phosphorus oxychloride. In organic solvents, organophosphorus condensation reagents tend to produce oligopeptides

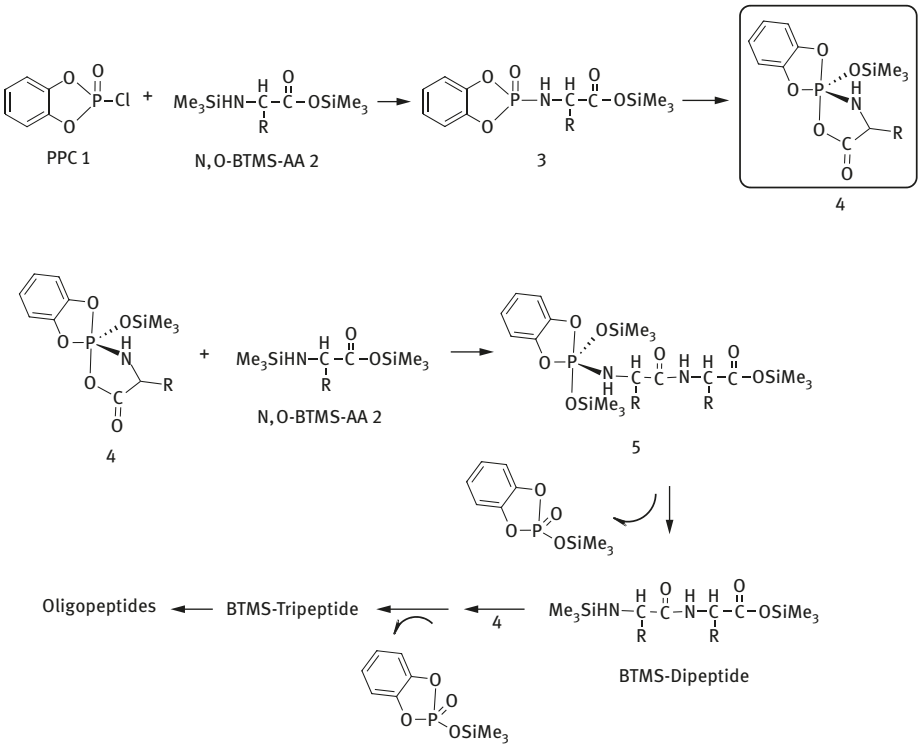


Figure 2.13: The possible mechanism of self-assembly into peptide for N,O-bis(trimethylsilyl)amino acids N,O-BTMS-AA mediated by PPC.

with high efficiency of amino acid polymerization. For example, the trimethylsilicon protected amino acids, such as glycine, phenylalanine, alanine, leucine and valine, can be polymerized into a peptide library with the treatment of phenyl phosphorus oxychloride (PPC). The five-membered cyclic pentacoordinate phosphorus intermediate 4 was formed in the PPC reaction with amino acid as shown in Figure 2.13. The intermediate 4 has also been successfully captured by using ^{31}P NMR spectrum with the chemical shift located in the typical five coordination phosphorus reagents approximately at −40 ppm. This intermediate 4 can be subsequently attacked by another molecule of protected amino acid to form dipeptides. The abovementioned reaction cycle eventually produces a variety of different polymerization degrees of peptide library. The formation of peptides using the organic phosphorus system provides key evidence for the process and mechanism of the peptide reaction of *N*-phosphoryl amino acids in the water system [19]. Interestingly, PPC reacts with silicon-etherified β-alanine or γ-butyric acid but cannot produce peptides. Although the chemical structure differences of amino acids is different merely with amino group positions, phosphorus reagents selectively distinguish structural differences of the amino acid and choose α-amino acid as the only substrate for the peptide in the process of condensation. This

process of amino acid selection has no enzymatic involvement, but is controlled only by the reactive nature of five-membered cyclic pentacoordinate phosphorus intermediate. The abovementioned mechanism provides the most favorable experimental evidence for underlining the possible role of phosphorus chemistry for selection of α-amino acids as basic unit for protein formation under prebiotic conditions.

2.6 Phosphorus chemistry and the molecular evolution of high-energy P–N bond

The prebiotic conditions are diverse, such as acid–base gradient, temperature gradient and metal ion catalysis. Under these prebiotic conditions, amino groups of amino acids are more easy to react with phosphorus reagents to form P–N bonds. The structure of P–N bond has good stability in alkaline conditions, but is very easy to cleave under acid conditions, indicating that the key process of reversible phosphorylation does not require the participation of enzyme and is simply only through the change of acid/alkali conditions to achieve the function of different molecules. In today's life system, carboxyl group of amino acids is easier to phosphorylate to form high-energy mixed anhydride structures with the catalysis of specific enzymes. These two different activation methods of amino acids at N-terminal amino group and C-terminal carboxylic group may imply some kind of evolutionary link between P–N and P–O bonds. Furthermore, the condensation process of amino acids to form amide bonds activated by phosphorus reagents under prebiotic conditions may amplify the chemical difference of α-amino acids with β– and γ-amino acids. The formation of five-membered cyclic pentacoordinate phosphate intermediates determined the selection of α-amino acids for peptide formations. In conclusion, the systematic regulation of phosphorus chemistry might play key roles in origin of the first proteins with biological functions and can be postulated that α-amino acids have to be chosen by phosphorus.

References

[1] Zhao YF, Zhao GH, Ma Y. Phosphorus and life sciences [M]. Beijing: Tsinghua University Press, 2005.
[2] Zimmer C. On the origin of life on earth [J]. *Science*, 2009, 323: 198–199.
[3] Westheimer, FH. Why nature chose phosphates [J]. *Science*, 1987, 235: 1173–1178.
[4] de Duve, C. Blueprint for a cell: The nature and origin of life. Patterson: New York, 1991.
[5] Weckwerth G, Schidlowski M. Phosphorus as a potential guide in the search for extinct life on mars [J]. *Adv. Space Res.*, 1995, 15: 185–191.
[6] Miller SL. A production of amino acids under possible primitive earth conditions [J]. *Science*, 1953, 117(3046): 528–529.
[7] Wang W Q, Kobayashi K, Ponnamperuma C. Prebiotic synthesis in a mixture of phosphine, methane, nitrogen and water-the possible role of phosphine in chemical evolution [J]. *Sci. Bull.*, 1985, 30: 281–284.

[8] Johnson AP, Cleaves HJ, Dworkin JP et al. The Miller volcanic spark discharge experiment [J]. *Science*, 2008, 322(5900): 404.
[9] Danger G, Plasson R, Pascal R. Pathways for the formation and evolution of peptides in prebiotic environments [J]. *Chem. Soc. Rev.*, 2012, 41: 5416–5429.
[10] Pascal R, Boiteau L, Commeyras A. From the prebiotic synthesis of α-amino acids towards a primitive translation apparatus for the synthesis of peptides [J]. *Top. Curr. Chem.*, 2005, 259: 69.
[11] Pasek MA. Rethinking early earth phosphorus geochemistry [J]. *Proc. Natl. Acad. Sci. USA*, 2008, 3: 853–858.
[12] Pasek MA, Block K. Lightning-induced reduction of phosphorus oxidation state [J]. *Nat. Geosci.*, 2009, 2: 553–556.
[13] Yamagata Y, Watanabe H, Saitoh M, Namba T. Volcanic production of polyphosphates and its relevance to prebiotic evolution [J]. *Nature*, 1991, 352: 516–519.
[14] Pasek MA, Kee TP, Bryant DE et al. Production of potentially prebiotic condensed phosphates by phosphorus redox chemistry [J]. *Angew. Chem. Int. Ed.*, 2008, 47: 7918–7920.
[15] Rabinowitz J, Flores J, Krebsbach R, Rogers G. Peptide formation in the presence of linear or cyclic polyphosphates [J]. *Nature*, 1969, 224: 795–796.
[16] Gao X, Liu Y, Xu PX, et al. α-Amino acid behaves differently from β- or γ-amino acids as treated by trimetaphosphate [J]. *Amino Acids*, 2008, 34: 47–53.
[17] Gao X, Ni F, Bao J, et al. Formation of cyclic acylphosphoramidates in mass spectra of N-monoalkyloxyphosphoryl amino acids using electrospray ionization tandem mass spectrometry [J]. *J. Mass Spectrom.*, 2010, 45: 779–787.
[18] Cheng CM, Liu XH, Li YM et al. N-Phosphoryl amino acids and biomolecular origins [J]. *Orig. Life Evol. Biosphere.*, 2004, 34: 455.
[19] Fu H, Li ZZ, Zhao YF,Tu GZ. Oligomerization of N, O-bis(trimethylsily)-α-amino acids into peptides mediated by O-phenylenephosphorochloridate [J]. *J. Am. Chem. Soc.*, 1999, 121: 291.

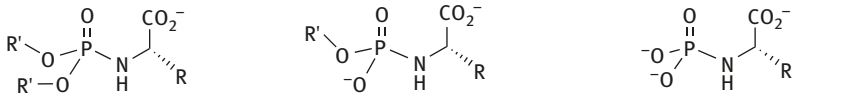
3 *N*-Phosphoryl amino acids – models for P–N bonds in prebiotic chemical evolution

Posttranslational modification of proteins with *N*-phosphorylation of basic amino acid residues plays important roles in biological processes. The high-energy P–N bond might have participated in the evolution of prebiotic chemistry. *N*-phosphoryl amino acids (PAAs) can serve as interesting small molecular models for the P–N bond in prebiotic chemical evolution. PAAs are capable of simultaneously producing several important biomolecules, such as polypeptides and oligonucleotides under mild reaction conditions. This chapter describes the chemistry of PAAs and discusses their likely prebiotic origin and their reactivity, and how they relate to biological P–N bond species. We depict a potentially prebiotic scenario mediated by PAAs where they act as one of the essential forces driving prebiotic biomolecules to the first protocell.

3.1 Introduction

Modifications of amino acid residues on proteins are essential events in biological processes. Being an important posttranslational protein modification, phosphorylation plays key roles in signal transduction, functional regulation and energy transfer [1, 2]. Phosphorylation on the side chain of a single amino acid residue (Ser, Thr and Tyr, forming a P–O ester bond, namely *O*-phosphorylation) on a protein may regulate its activity or make it recognizable by other biomacromolecules (e.g., DNA, RNA or protein). Phosphorylation of acidic (Asp, Glu, forming a P–O anhydride bond) and basic (His, Arg, Lys, forming a P–N bond, namely *N*-phosphorylation) amino acid residues normally generates high-energy species that are thought to be intermediates for enzymatic catalysis [3]; the main research works of protein phosphorylation and their biological functions are limited to Ser, Thr and Tyr [4]. More specifically, phosphorylation of hydroxyl (–OH) groups releases –6 to –9 kcal/mol Gibbs free energy, compared to –12 to –14 kcal/mol produced by phosphorylation of =NH groups [5]. In addition, biological P–N bond containing species are acid labile and are readily susceptible to spontaneous hydrolysis resulting in false negative by conventional analysis involving acidic treatments. Hence, our appreciation for P–N bond species has developed later than that of its P–O bond counterpart [6, 7]. The importance of biological P–N species has been recognized as being comparable to that of P–O bond species in cell functioning [4]. Recently, there has been significant progress in the design and development of amino acid phosphoramidates with P–N bonds as an effective prodrug strategy (e.g., sofosbuvir) [8, 9]. From a top-down view, one may ask: “Is the biological P-N bond species reminiscent of prebiotic chemistry evolution?”

https://doi.org/10.1515/9783110562552-003



N-Dialkyloxyphosphoryl amino acids (NDAPAAs)

N-Monoalkyloxyphosphoryl amino acids (NMAPAAs)

N-phosphono-amino acids or N-phospho-amino acids (NPAAs)

Figure 3.1: Structures and nomenclature of *N*-phosphoryl amino acids.

After systematically studying three categories of *N*-phosphoryl amino acids (PAAs, Figure 3.1), we proposed that PAAs can serve as interesting models for the P–N bond in prebiotic chemical evolution [10, 11], and some pivotal biomolecules, including peptide, oligonucleotide, phosphoester, phosphoramidate, aminoacyl phosphate and nucleoside mono/diphosphates can be generated from PAA-mediated reactions under mild conditions. This tutorial review is an updated version to our previous review on the chemistry of *N*-dialkyloxyphosphoryl amino acids (NDAPAAs) [10], and will cover the chemistry of *N*-monoalkyloxyphosphoryl amino acids (NMAPPAAs) and *N*-phosphono-amino acids (NPAAs). We will discuss their likely prebiotic origin, reactivity and potential link to biological P–N bond species.

3.2 The P–N bond in modern biology

At least nine phosphorylated amino acids (serine, threonine, tyrosine, histidine, lysine, arginine, aspartate, glutamate and cysteine) have been identified in biological systems [7]. Since acid treatment is routine during purification and analysis of phosphoproteins, the study of phosphorylated basic amino acids in proteins and corresponding protein kinases has made slower progress than that of Ser, Thr and Tyr. However, contributions of *N*-phosphorylation in protein to cell signaling pathway and relevant physiological regulation should not be underestimated. For instance, the role of histidine phosphorylation and the two-component histidine kinases in bacterial sensor signaling was recognized in the 1990s and also analogous systems were later found in plants and fungi [12, 13]. In the 1970s, Smith and coworkers reported the occurrence of histone H4 histidine kinase (HHK) in the nuclei of rat hepatocytes, which is associated with cell proliferation and differentiation [14, 15]. Moreover, histidine phosphorylation of P-selectin upon stimulation of human platelets was identified as a novel signal transduction pathway [16]. Nowadays we know that protein *N*-phosphorylation species exist in the form of phosphohistidine, phosphoarginine and phospholysine (Figure 3.2). In general, these phosphorylated proteins serve as either high-energy species for phosphoryl group transferring, or signal transduction components [4, 7, 17].

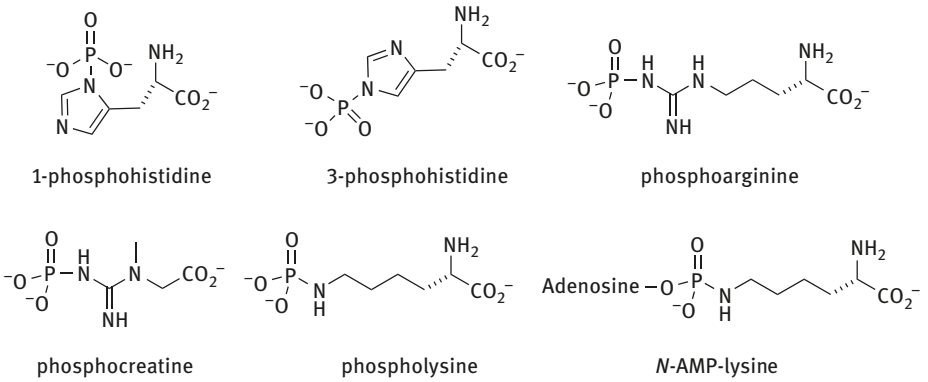


Figure 3.2: Examples of biological P–N bond species.

3.2.1 Phosphohistidine

Covalent binding of phosphate to histidine residues (pHis) was first detected in 1962 [18]. Followed by decades of silence, the importance of pHis in vertebrates is being recognized. The key finding in this field is the discovery of a pHis phosphatase [19]. This immediately stimulated an investigation of the biological implications of pHis in mammals. As the most studied *N*-phosphorylation form of amino acid residue, though not that much, there are two biologically relevant pHis isomers: (1) N^1-phosphohistidine (π-phosphohistidine) and (2) N^3-phosphohistidine (τ-phosphohistidine) (Figure 3.2) that have been identified among prokaryotes, lower eukaryotes and mammals [7]. Besides as a catalytic intermediate in nucleoside diphosphate kinase (NDPK) that catalyzes the interconversion of nucleoside diphosphates and triphosphates [20], there are three systems where pHis residue has a substantial impact on cellular function and metabolism: the β subunit of heterotrimeric guanosine triphosphate (GTP)-binding proteins (G proteins) [21–25], the enzyme adenosine 5′-triphosphate (ATP)-citrate lyase (ACL) [26–29] and the Ca^{2+}-activated K^+ channel KCa3.1 [30, 31].

3.2.2 Phosphoarginine and phosphoryl-lysine

Despite the widely known modifications of arginine and lysine (e.g., methylation, acetylation, citrullination, SUMOylation and ubiquitination) and their role in epigenetics and ubiquitin-proteasome system [32, 33], protein Arg and Lys phosphorylation (pArg and pLys) has been studied less widely, not to say their role in signal transduction. Beside experimental difficulties in analysis and peptide synthesis [34, 35], pArg and pLys appear to be present much less frequently than pHis [3]. Wakim et al. have showed that histone H3 undergoes phosphorylation by a

Ca^{2+}-calmodulin-dependent kinase on four arginine residues, three of them located within the C-terminal, resulting in alteration of overall charge of that region and potentially regulating the binding of H3 to DNA during nucleosome assembly/disassembly [36]. Nevertheless, the discovery of a bacterial arginine kinase (McsB) that phosphorylates a heat-shock regulator provided the first example of a pArg-based signaling mechanism [37]. This well-characterized system complements previous sporadic reports and has definitely evidenced the natural and *in vivo* occurrence of pArg [17, 36, 38–40].

Biological pLys was first identified in 1967 when Zetterqvist isolated *N*-ε-^{32}P-phospholysine along with N^1-^{32}P-phosphohistidine, and N^3-^{32}P-phosphohistidine from an alkaline hydrolysate of bovine liver proteins incubated with γ-^{32}P-ATP [41]. From then on, a series of experiments were reported by Smiths and coworkers along with their studies on histidine kinase [42]. Till now, pLys is believed to be present in nuclei of rat liver, myosin and histones of some other cells [42–44].

3.3 Potential prebiotic origins of *N*-α-phosphoryl amino acids

3.3.1 The phosphorus problem

Prebiotic formation of PAAs requires adequate amino acids and phosphorus compounds on the primitive Earth. The likely origin of amino acids is no longer a problem. There are two clear sources of amino acids in the primitive Earth: [45, 46] (1) it was exogenous when amino acids were synthesized outside the Earth and delivered to our planet by interplanetary dust particles, meteorites, comets and so on and (2) it was endogenous when amino acids were synthesized on the Earth in atmospheric mixtures, hydrothermal vents and so on.

Phosphorus is no doubt an essential element for life and presumably plays critical role in the origin of life [47–49]. In 1981, Nobel Prize winner Alexander Todd said "If life exists anywhere else in the universe it will do so only on a planet on which phosphorus is readily available" [50]. De Duve was completely right when he stated that "Life is basically organized around phosphorus" [51]. However, prebiotic availability of phosphate has long been questioned in terrestrial abiogenesis and chemical evolution [52]. Despite with low solubility and reactivity, the mineral apatite has long been considered as the only significant source of phosphate on Earth. It has been also the topic of much debate on the availability of phosphates and polyphosphates in aqueous solution and being precipitated by metal cations [53].

To answer the "phosphorous problem", Schwartz had given a detailed account of phosphorus sources on the Earth, such as schreibersite, phosphate, hypophosphite and polyphosphates [49, 54], and stated that "phosphorus problem" is no

longer the stumbling block that it was once thought to be. Recently, laboratory measurements showed that Martian phosphate minerals dissolve 45 times more quickly than terrestrial ones, resulting in more than twice phosphate concentration compared to those on Earth. This result may have mitigated one of the hurdles to abiogenesis on Mars and gave another answer to the "phosphorus problem" [55]. Given the opportunity to react with amino acids, phosphate can be further solubilized by forming N^{α}-phosphoryl amino acids on the primitive Earth. Although all abovementioned studies can only provide a glimpse in the origin of life, it is not imprudent to believe that such kind of reaction might happen in localized "ponds" or "lagoons" on primitive Earth.

In this chapter, we would like to emphasize that P_3m (trimetaphosphate) proposed as a prebiotic phosphorus source might be exempted from "phosphorus problem" [56], because P_3m in the form of $Ba_3(P_3O_9)_2$ and $Mg_3(P_3O_9)_2 \cdot 13H_2O$ are easily soluble in water (Gmelin database), suggesting that P_3m can survive in aqueous solution in the presence of divalent metal ion. On account of its unique properties, P_3m has been subjected to chemical evolution studies, such as the synthesis of glycolaldehyde phosphate [57] and amphiphiles [58]. Interestingly, in our attempt to isolate NPAAs [59] from unreacted P_3m and by-products, pyrophosphate and triphosphate are all precipitated by adding calcium oxide (CaO) but leaving P_3m intact. This result implies that pyrophosphate and triphosphate might be sacrificed for protecting P_3m from precipitation.

3.3.2 Possible pathways for origin of *N*-phosphoryl amino acids

NPAAs are the simplest PAAs lacking phosphate ester groups. Formation of NPAAs under potential prebiotic environment may involve reaction of amino acids with polyphosphate, especially trimetaphosphate (P_3m) [60]. Attacking of P_3m by the α-amino group of amino acids leads to a Triphosphoryl-amino acid (P_3-AA) intermediate followed by fragmentation into cyclic acyl phosphoramidate (CAPA2) and pyrophosphate. Since CAPA2 is unstable, it is subsequently hydrolyzed to NPAA with ring opening (Figure 3.3) [61]. Recently, the above-mentioned process and CAPA2 have been trapped in gas phase by electrospray ionization tandem mass spectrometry [62]. Based on the abovementioned mechanism, a green synthetic method was also developed for preparation of NPAAs on gram scale [59].

NMAPAAs have one more ester group on phosphate than NPAAs. To synthesize NMAPAAs, the phosphorus precursor should possess two substitutable leaving groups (e.g., P_3m) or two coupling sites (e.g., hypophosphite). The first plausible pathway to NMAPAAs is through attacking of alkoxyl nucleophile (ROH) on intermediate CAPA2 generated in the reaction of P_3m with amino acids. (Figure 3.3) [61]. For instance, *N*-mono-methoxyphosphoryl glycine was detected in a mixture of glycine (0.1 M) and P_3m (0.1 M) in methanol/water (v: v=1:4) solution.

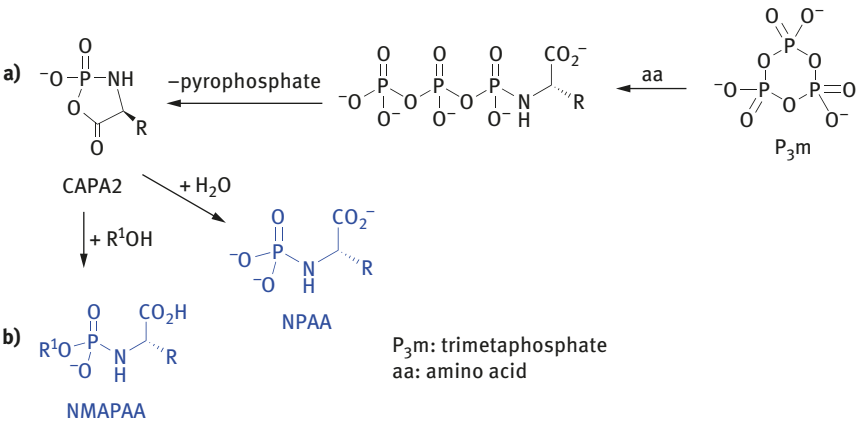


Figure 3.3: Potential pathways of PAA formation mediated by various prebiotic phosphorus sources.

3.4 Reactivity of N^{α}-phosphoryl amino acids related to prebiotic chemistry

3.4.1 *N*-dialkyloxyphosphoryl amino acids

Although prebiotic formation path for NDAPAAs is still hypothetical, NDAPAAs have been studied extensively to introduce a series of biomimetic reactions (Figure 3.4) [8]. It was found that when amino acids were phosphorylated, peptide formation [63], ester formation [64], ester exchange at phosphorus [65], N to O [66] and N to S migration [67] and oligonucleotide formation [68] were observed at room temperature (Figure 3.4). The phosphorus atom is the center; carboxyl group and side chain group participate in the reaction at the same time, which is similar to the arrangement of active site in enzymes. Therefore, we considered *N*-phosphoryl amino acid and small peptide as "mini activating enzymes." The remarkable reaction of *N*-phosphoryl amino acids with nucleoside mixtures, forming peptides and oligonucleotides simultaneously, suggests a pathway of "co-evolution of proteins and nucleic acids" [68, 69]. We presumed that a pentacoordinate phosphorus intermediate was involved in the reaction processes. However, the intermediate would be unstable in water and alcohol and it might only exist as a transient intermediate. To prove the existence of such a pentacoordinate structure in the reaction mechanism, the *N*,*O*-bis(trimethylsilyl) protected α-amino acids, such as alanine, valine and phenylalanine, were reacted with *O*-phenylene phosphorochloridate to yield an activated α-amino acid in the form of a five-membered cyclic pentacoordinate phosphoric-carboxylic mixed anhydride (P5-CAPA1) that could be observed by ^{31}P-NMR [70]. However, under the same conditions, no P5-CAPA1 was detected for β-amino acids. Furthermore, several stable bisamino acyl pentacoordinate spirophosphoranes derived from α-amino acids were

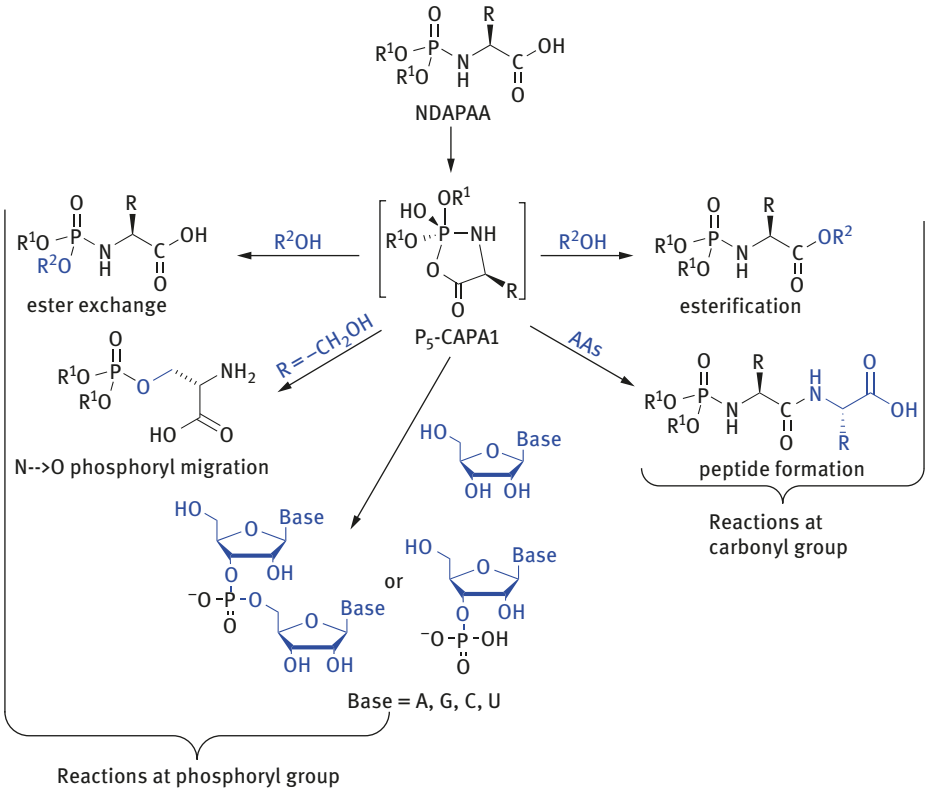


Figure 3.4: NDAPAA-mediated formation of biomolecules. All pathways have been experimentally supported.

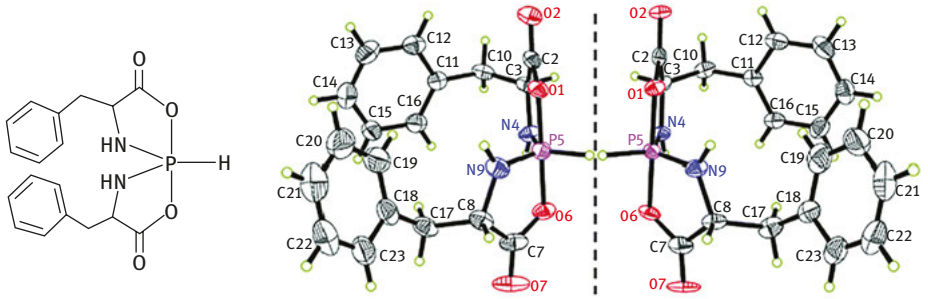


Figure 3.5: The crystal structures of two stereoisomers of pentacoordinate spirophosphoranes derived from phenylalanine [71].

synthesized and their absolute configurations were determined by single crystal X-ray diffraction, nuclear magnetic resonance (NMR), circular dichroism (CD) and theoretical calculations (Figure 3.5) [71, 72]. The versatility of NDAPAAs is further complicated

by attack of nucleophiles on both carbonyl and phosphoryl sites of P5-CAPA1 [61]. *N*-phosphoryl-peptide formation and esterification belong to carbonyl-directed reactions in ester exchange; migration of N to O phosphoryl and formation of oligonucleotide result from phosphoryl-directed reactions.

Why nature chose α-amino acids is still a mystery. After studying the reactivity of *N*-phosphoryl β- and γ-amino acids, we found that they did not show peptide formation [70, 73] when the same conditions of *N*-phosphoryl α-amino acid reactions were applied. To understand the different reactivities of *N*-phosphoryl-α-amino acids and *N*-phosphoryl β- or γ-amino acids (Figure 3.6), *N*-dimethylphosphorylaspartic acid (NDMPAA) [74] with a free β-COOH group in the side chain and *N*-dimethylphosphorylglutamic acid (NDMPGA) [75] with a free γ-COOH were chosen as model reactants for in silico generation of pentacoordinate phosphorus intermediates IN1, IN2, IN3 and IN4 (Figure 3.7). At the B3LYP/6-31G** level, natural bond orbital (NBO) analysis showed that positive charge of the proton in α-COOH is about 0.0027 larger than that of the proton in the β-COOH. Meanwhile, the β-$CO2^-$ anion derived from NDMPAA is less stable than the α-$CO2^-$ anion by 18.51 kJ/mol at the HF/6-31G** level, and the β-$CO2^-$ anion would be transferred to the α-$CO2^-$ anion at the B3LYP/6-31G** level.

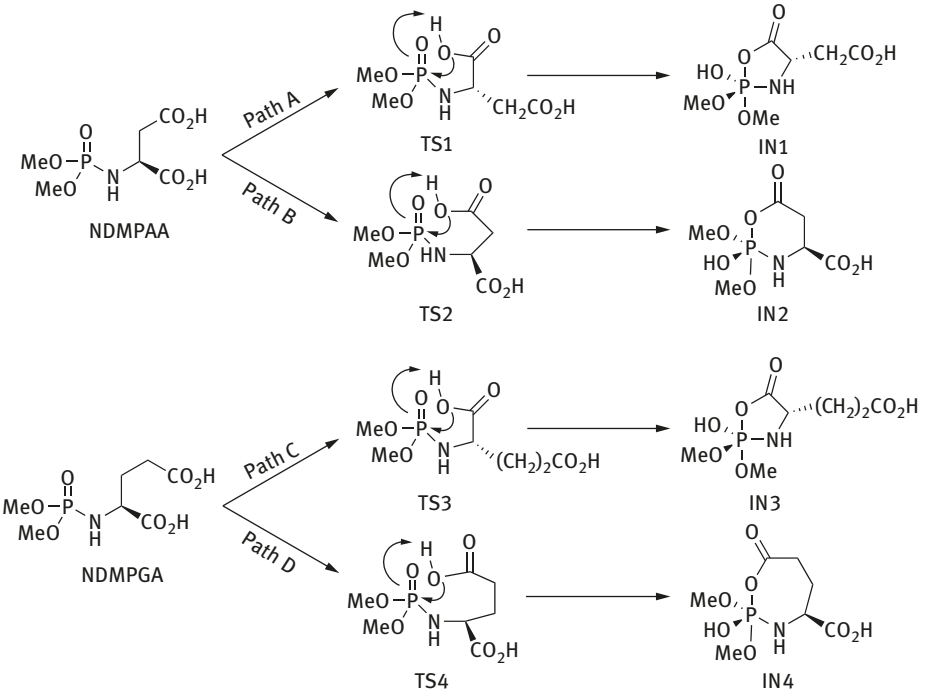


Figure 3.6: Possible reaction mechanism involved in formation of pentacoordinate phosphorane intermediates from *N*-phosphoryl amino acids.

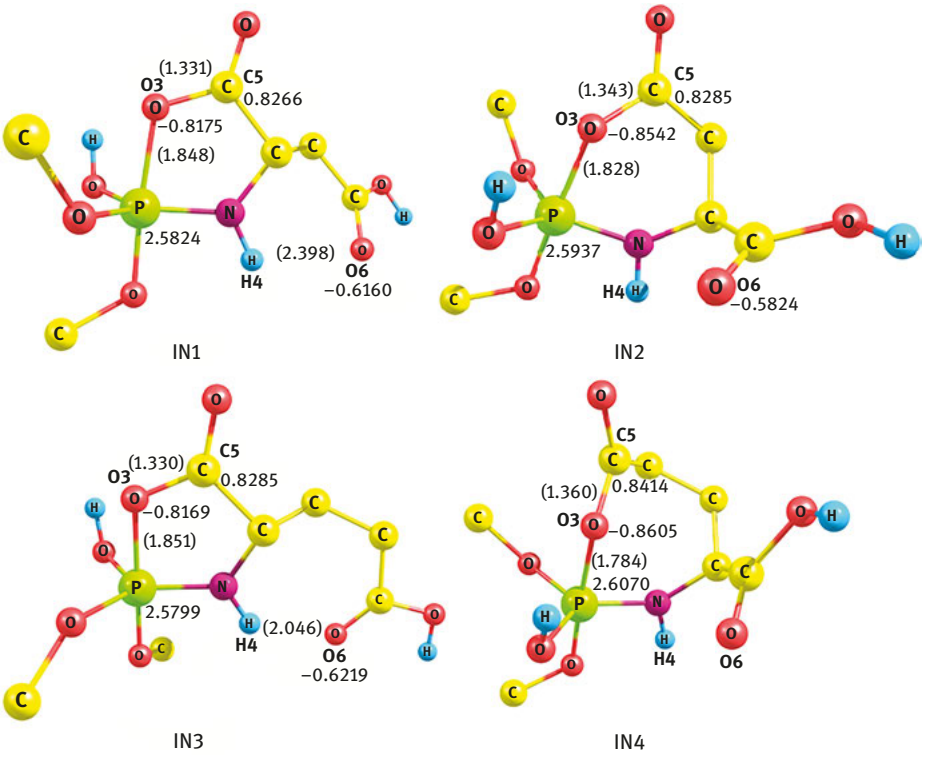


Figure 3.7: Optimized structures of In1-4 at B3lyp/6-31g** level, showing NBO atomic charges and distances (Unit in Parentheses, Å).

Therefore, a proton might be easier to transfer from an α-COOH than from a β-COOH, favoring the nucleophilic attack of α-COOH group on the phosphorus atom. Meanwhile, two transition states TS1 and TS2 involving an α-COOH or β-COOH group were identified (Figure 3.6). The energy of TS2 is calculated to be 45.74 kJ/mol higher than that of TS1. The same study on NDMPGA revealed that IN4 was 66.56 kJ/mol higher in energy than IN3 at the B3LYP/6-31G** level and the transition states TS3 and TS4 involving an α-COOH or γ-COOH group had energy barriers of ΔE = 57.59 kJ/mol and 120.93 kJ/mol, respectively. As a result, formation of P5-CAPA1 favors NDAPAAs contains an α-COOH group.

3.4.2 Reactivity and Potential Prebiotic Chemistry of NMAPAAs

Compared to NDAPAAs, NMAPAAs are relatively stable at neutral pH. This might be explained by unfavorable electrostatic repulsion between the negatively charged phosphoryl group and carboxyl group during intramolecular cyclization. However,

NMAPAAs are reactive in weakly acidic aqueous solution (pH = 4.0–5.0). A series of products were identified in the reactions of model compounds *N*-monomethoxyphosphoryl glycine, alanine and valine [11]. The proposed mechanism is interesting. Protonation of carboxylate and phosphoramidate nitrogen at a weak acidic pH leads to two reactive intermediates called NMAPAA-H^O and NMAPAA-H^N, respectively (Figure 3.8).

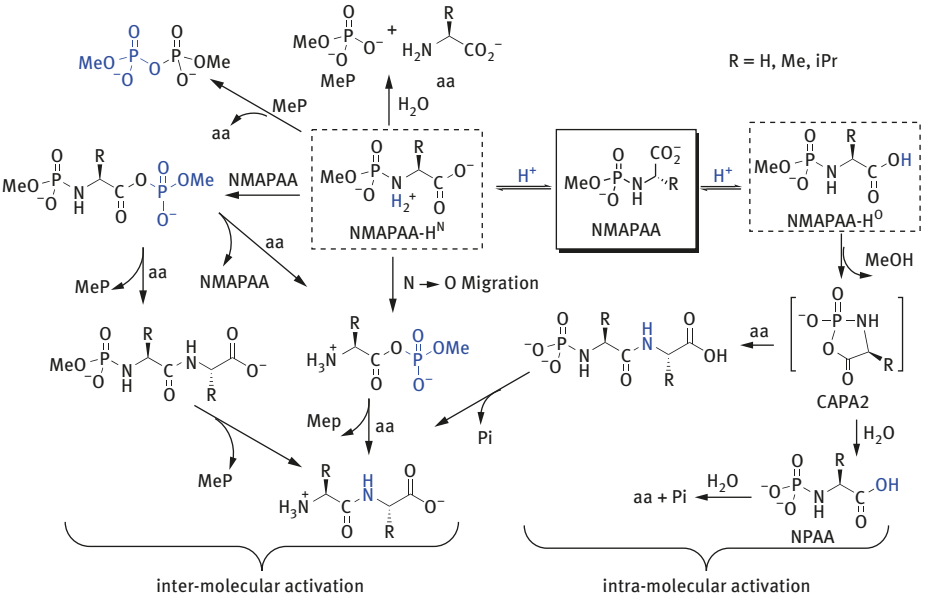


Figure 3.8: NMAPAA-mediated formation of biomolecules.

NMAPAA-H^O is believed to follow an intramolecular cyclization and thus activates NMAPAAs into the cyclic intermediate CAPA2, and then gives *N*-phosphopeptide and NPAA. While protonation of nitrogen (forming NMAPAA-H^N) will weaken the P–N bond and make the amino acid a good leaving group for nucleophilic attack of various species. Introduction of Mg^{2+} into the reactions increases the yield of dimethyl pyrophosphate but decreases the peptide yield (Table 3.1, entries 4–6). NMAPAAs are not only self-activated to produce "high-energy" phosphoric-carboxylic anhydrides, but also deliver reasonable yields of oligopeptides for several NMAPAAs under various reaction conditions (Table 3.1, [11]). Variations in the chemical structure of the alkyl group and the concentration of *N*–MMP–Gly had no significant effect on the yields of phosphates and peptides (entries 1–3). However, the product yields depend on the structures of amino acid moieties of NMAPAAs. In the case of N–MMP–Val (entry 8), the yield of dipeptide Val–Val is low (1.0%), whereas the yield of phosphate is as high as 68.2%. This finding can be rationalized in terms of the steric effect of the hydrophobic side

chains of the amino acids or the so-called near attack conformer effect [76]. Larger side chains will compress the carboxylic acid and phosphoryl groups together and consequently facilitate nucleophilic attack of the carboxylate on the phosphorus with expulsion of methanol to form a cyclic acylphosphoramidate (CAPA2), which favors hydrolysis and finally produces substantial amount of Pi (entries 7–8) instead of producing MeP in the Gly analogue.

Table 3.1: Yields of phosphates and peptides produced from NMAPAAs [11].

Entry	Reactions[a]	Yield of phosphate [%][b]			Yield of peptide [%][c]
		MeP	Pi	P_2	
1	MMP–Gly (0.2)[d]	89.7	5.1	5.2	G_2 (13.0), G3 (1.2), G4 (0.1)
2	MMP–Gly (0.5)	88.4	5.0	6.6	G_2 (16.0), G3 (1.7), G4 (0.2)
3	MEP–Gly (0.2)[d]	91.8	3.0	5.2	G_2 (11.5), G3 (1.0)
4	MMP–Gly (0.2) + Mg^{2+} (0.02)	88	5.7	6.3	G_2 (6.4), G3 (0.3)
5	MMP–Gly (0.2) + Mg^{2+} (0.05)	84.1	4.1	11.8	G_2 (6.8), G3 (0.3)
6	MMP–Gly (0.2) + Mg^{2+} (0.1)	78.4	4.4	17.2	G_2 (7.9), G3 (0.4)
7	MMP–Ala (0.2)	53.9	40.4	5.7	A_2 (4.6), A3 (0.2)
8	MMP–Val (0.2)	23.5	68.2	8.3	V_2 (1.0)
9	MMP–Gly (0.2) + Ala (0.2)	90.7	5.7	3.6	G_2 (3.8), GA (4.0)[e], AG (3.5), A_2 (3.9)

[a]Reaction conditions: incubation at 40°C, pD = 4.0–5.0, reaction time: entries 1, 3–6 and 9: 8 d; entry 7: 10 d; entries 2 and 8: 12 d.
[b]Yields were determined by integration of the 31P NMR spectra. MeP: methyl phosphate; Pi: PO_4^{3-}; P2: dimethyl pyrophosphate.
[c]Peptide yields were determined relative to *N*-phosphoryl amino acids by HPLC.
[d]MMP–Gly: sodium salt of *N*-(methoxyphosphoryl) glycine; MEP–Gly: sodium salt of *N*-(ethoxyphosphoryl) glycine.
[e]Including a small amount of triglycine.

3.4.3 *N*-phosphono-amino acids

NPAAs are nonesterified phosphoramidate with pKa values ~8.5 (at nitrogen atom). Therefore, the nitrogen atom of NPAAs is fully protonated at neutral pH, which weakens the P–N bond and makes NPAAs good phosphate donors. *N*-phosphono-Gly and *N*-phosphono-Ala have proved active in both peptide formation and phosphate transfer (Figure 3.9). These processes involve second-order reactions for concentrated NPAA solutions near pH = 7–8 at moderate temperature. Phosphate transfer between two NPAAs leads to *N*-pyrophospho-amino acids that spontaneously produce CAPA2 for peptide formation. Under Mg^{2+} catalysis, nucleoside monophosphates (AMP, GMP, CMP and UMP) were phosphorylated to give nucleoside diphosphates by NPAAs.

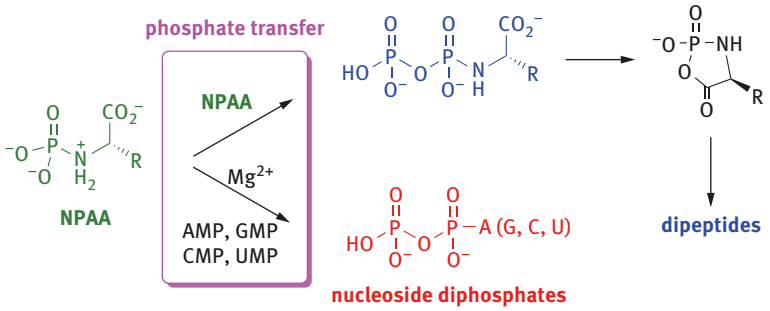


Figure 3.9: NPAA -mediated formation of biomolecules.

Although NPAAs are potential molecules with the prebiotic significance shown in Figure 3.9, they are not stable around neutral pH. The half-life of α-NPAAs is only 3.6–18 h at pH = 7 in water, but it is relatively stable at pH 10–13, which can survive under extreme alkaline conditions [77].

3.5 Summary

There have been many theoretical and model systems proposed for the origin of biomolecules. However, as a model system, the PAAs are capable of simultaneously producing several important biomolecules. So this model system is unique and novel. We believe that the origin of several biomolecules at the same "location" would greatly accelerate their assembly into advanced and functional complexes. Here we depict a prebiotic scenario mediated by PAAs. In this scenario, PAAs act as one of the essential forces driving prebiotic biomolecules to the first protocell (Figure 3.10). When accumulated active phosphorus reacted with amino acids on the primitive Earth, *N*-phosphoryl amino acids formed and began to direct their reactions toward the synthesis of biomolecules, now recognized as key components of contemporary cells [78]. These biomolecules include peptide/proteins, oligonucletide/RNAs, phosphoester/lipids, phosphoramidates, aminoacyl phosphates and nucleoside mono/di/triphosphates. Given appropriate conditions (inclusion of metal ions), these biomolecules can assemble into protocells. As energy and information carriers, PAAs might have existed in protocells for a limited period before more advanced cells emerged. Then, P–N species, in the form of pHis, pArg and pLys proteins in contemporary cells, might have inherited some of the reactivity of PAAs during their prebiotic chemistry evolution. These events, involving a high-energy P–N bond as energy transfer intermediates, such as the synthesis of carboxyl-phosphate anhydride and phosphoanhydride, were later taken over by contemporary enzymes such as histidine kinase and NDPK, respectively. There is another possible control mechanism, given

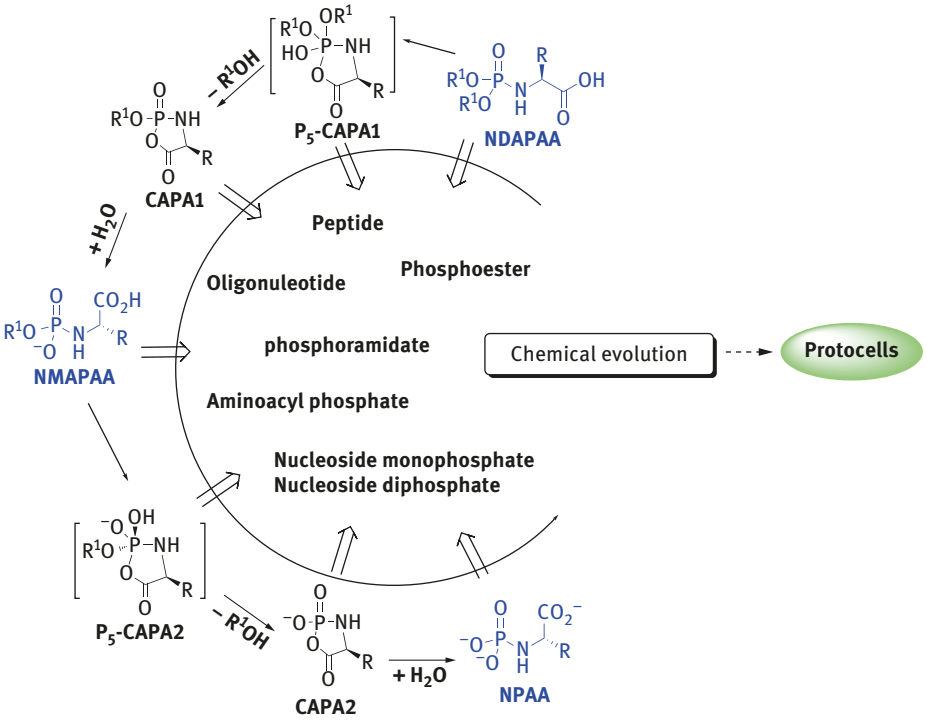


Figure 3.10: Metabolic pathways of PAAs related to chemical evolution.
Note: From *N*-phosphoryl amino acids to biological P–N bond species.

a favorable acidic pH in the intracellular microenvironment, the acid-labile nature of the P–N bond of pHis may be employed in a system that requires a pH-sensitive on/off switch without the need for additional protein phosphatases [7]. Similar chemistry is observed for NMAPAAs and NPAAs that become labile at acidic pH. In summary, we hypothesize that the high-energy P–N bond might have aggressively participated in the evolution of prebiotic chemistry, possibly through carrier PAAs. In contemporary biology, some cellular processes are still regulated *via* *N*-phosphorylated proteins, which can be regarded as successors to PAAs in prebiotic chemical evolution.

3.5.1 Perspectives

PAAs might have played an important role in the prebiotic chemical evolution. This system deserves further development in two directions. One direction is to identify more and easier accessible pathways for the prebiotic synthesis of PAAs. Exploring the basic molecular recognition pattern between amino acids and nucleotides may

reveal some clues to the origin of genetic code and chirality that PAAs provide for an ideal scaffolding of molecular interactions between amino acids and nucleotides via the linkage on phosphate moiety [79, 80].

References

[1] Derouiche A, Cousin C, Mijakovic I. Protein phosphorylation from the perspective of systems biology [J]. *Curr Opin Biotechnol*, 2012, 23:585–590.
[2] Hunter T. Why nature chose phosphate to modify proteins [J]. *Phil Trans R Soc B*, 2012, 367:2513–2516.
[3] Matthews HR. Protein-kinases and phosphatases that act on histidine, lysine, or arginine residues in eukaryotic proteins – a possible regulator of the mitogen-activated protein-kinase cascade [J]. *Pharmacol Ther*, 1995, 67:323–350.
[4] Ciesla J, Fraczyk T, Rode W. Phosphorylation of basic amino acid residues in proteins: important but easily missed [J]. *Acta Biochim Pol*, 2011, 58:137–148.
[5] Zu XL. *Wilson and Wilson's Comprehensive Analytical Chemistry* [M]. London: Elsevier, 1959, l–52.
[6] Loroch S, Dickhut C, Zahedi RP, Sickmann A. Phosphoproteomics--more than meets the eye [J]. *Electrophoresis*, 2013, 34: 1483–1492.
[7] Attwood PV, Piggott MJ, Zu XL, Besant PG. Focus on phosphohistidine [J]. *Amino Acids*, 2007, 32:145–156.
[8] Serpi M, Bibbo R, Rat S, Roberts H, Hughes C, Caterson B, Jose Alcaraz M, Torrent Gibert A, Alaez Verson CR, McGuigan C. Novel phosphoramidate prodrugs of N-acetyl-(D)-glucosamine with antidegenerative activity on bovine and human cartilage explants [J]. *J Med Chem*, 2012, 55:4629–4639.
[9] Reardon S. United States to approve potent oral drugs for hepatitis C [J]. *Nature*, 30 October 2013.
[10] Cheng CM, Liu XH, Li YM, Ma Y, Tan B, Wan R, Zhao YF. N-phosphoryl amino acids and biomolecular origins [J]. *Orig Life Evol Biosph*, 2004, 34:455–464.
[11] Gao X, Deng H, Tang G, Liu Y, Xu P, Zhao Y. Intermolecular phosphoryl transfer of N-phosphoryl amino acids [J]. *European J Org Chem*, 2011:3220–3228.
[12] Hess JF, Bourret RB, Simon MI. Histidine phosphorylation and phosphoryl group transfer in bacterial chemotaxis [J]. *Nature*, 1988, 336:139–143.
[13] Swanson RV, Alex LA, Simon MI. Histidine and aspartate phosphorylation – 2-component systems and the limits of homolog [J]. *Trends Biochem Sci*, 1994, 19:485–490.
[14] Besant PG, Attwood PV. Histone H4 histidine phosphorylation: Kinases, phosphatases, liver regeneration and cancer [J]. *Biochem Soc Trans*, 2012, 40:290–293.
[15] Smtih DL, Bruegger BB, Halpern RM, Smith RA. New histone kinases in nuclei of rat tissues [J]. *Nature*, 1973, 246:103–104.
[16] Crovello CS, Furie BC, Furie B. Histidine phosphorylation of P-selectin upon stimulation of human platelets: A novel pathway for activation-dependent signal transduction [J]. *Cell*, 1995, 82:279–286.
[17] Besant PG, Attwood PV, Piggott MJ. Focus on Phosphoarginine and Phospholysine [J]. *Curr Protein Pept Sci*, 2009, 10:536–550.
[18] Boyer PD, Deluca M, Ebner KE, Hultquist DE, Peter JB. Identification of phosphohistidine in digests from a probable intermediate of oxidative phosphorylation [J]. *J Biol Chem*, 1962, 237:PC3306–PC3308.

[19] Ek P, Pettersson G, Ek B, Gong F, Li JP, Zetterqvist O. Identification and characterization of a mammalian 14-kDa phosphohistidine phosphatase [J]. *Eur J Biochem*, 2002, 269:5016–5023.
[20] Munoz-Dorado J, Almaula N, Inouye S, Inouye M. Autophosphorylation of nucleoside diphosphate kinase from Myxococcus xanthus [J]. *J Bacteriol*, 1993, 175:1176–1181.
[21] Wieland T, Nurnberg B, Ulibarri I, Kaldenbergstasch S, Schultz G, Jakobs KH. Guanine nucleotide-specific phosphate transfer by guanine-nucleotide-binding regulatory protein beta-subunits – characterization of the phosphorylated amino-acid [J]. *J Biol Chem*, 1993, 268:18111–18118.
[22] Kowluru A, Seavey SE, Rhodes CJ, Metz SA. A novel regulatory mechanism for trimeric GTP-binding proteins in the membrane and secretory granule fractions or human and rodent beta cells [J]. *Biochem J*, 1996, 313:97–107.
[23] Cuello F, Schulze RA, Heemeyer F, Meyer HE, Lutz S, Jakobs KH, Niroomand F, Wieland T. Activation of heterotrimeric G proteins by a high energy phosphate transfer via nucleoside diphosphate kinase (NDPK) B and Gbeta subunits. Complex formation of NDPK B with Gbeta gamma dimers and phosphorylation of His-266 IN Gbeta [J]. *J Biol Chem*, 2003, 278:7220–7226.
[24] Maurer A, Wieland T, Meissl F, Niroomand F, Mehringer R, Krieglstein J, Klumpp S. The beta-subunit of G proteins is a substrate of protein histidine phosphatase [J]. *Biochem Biophys Res Commun*, 2005, 334:1115–1120.
[25] Hippe HJ, Lutz S, Cuello F, Knorr K, Vogt A, Jakobs KH, Wieland T, Niroomand F. Activation of heterotrimeric G proteins by a high energy phosphate transfer via nucleoside diphosphate kinase (NDPK) B and Gbeta subunits. Specific activation of Gsalpha by an NDPK B.Gbetagamma complex in H10 cells [J]. *J Biol Chem*, 2003, 278:7227–7233.
[26] Wagner PD, Vu ND. Phosphorylation of atp-citrate lyase by nucleoside diphosphate kinase [J]. *J Biol Chem*, 1995, 270:21758–21764.
[27] Wagner PD, Steeg PS, Vu ND. Two-component kinase-like activity of nm23 correlates with its motility-suppressing activity [J]. *Proc Natl Acad Sci USA*, 1997, 94:9000–9005.
[28] Klumpp S, Bechmann G, Maurer A, Selke D, Krieglstein J. ATP-citrate lyase as a substrate of protein histidine phosphatase in vertebrates [J]. *Biochem Biophys Res Commun*, 2003, 306:110–115.
[29] Krieglstein J, Lehmann M, Maurer A, Gudermann T, Pinkenburg O, Wieland T, Litterscheid S, Klumpp S. Reduced viability of neuronal cells after overexpression of protein histidine phosphatase [J]. *Neurochem Int*, 2008, 53:132–136.
[30] Srivastava S, Li Z, Ko K, Choudhury P, Albaqumi M, Johnson AK, Yan Y, Backer JM, Unutmaz D, Coetzee WA, Skolnik EY. Histidine phosphorylation of the potassium channel KCa3.1 by nucleoside diphosphate kinase B is required for activation of KCa3.1 and CD4 T cells [J]. *Mol Cell*, 2006, 24:665–675.
[31] Srivastava S, Zhdanova O, Di L, Li Z, Albaqumi M, Wulff H, Skolnik EY. Protein histidine phosphatase 1 negatively regulates CD4 T cells by inhibiting the K(+) channel KCa3.1 [J]. *Proc Natl Acad Sci USA*, 2008, 105:14442–14446.
[32] Freiman RN, Tjian R. Regulating the regulators: Lysine modifications make their mark [J]. *Cell*, 2003, 112:11–17.
[33] Smith BC, Denu JM. Chemical mechanisms of histone lysine and arginine modifications [J]. *Biochimica et Biophysica Acta*, 2009, 1789:45–57.
[34] Hofmann FT, Lindemann C, Salia H, Adamitzki P, Karanicolas J, Seebeck FP. A phosphoarginine containing peptide as an artificial SH2 ligand [J]. *Chem Commun (Camb)*, 2011, 47:10335–10337.
[35] Bertran-Vicente J, Serwa RA, Schumann M, Schmieder P, Krause E, Hackenberger CP. Site-specifically phosphorylated lysine peptides [J]. *J Am Chem Soc*, 2014, 136:13622–13628.

[36] Wakim BT, Aswad GD. Ca2+-calmodulin-dependent phosphorylation of arginine in histone-3 by a nuclear kinase from mouse leukemia-cells [J]. *J Biol Chem*, 1994, 269:2722–2727.
[37] Fuhrmann J, Schmidt A, Spiess S, Lehner A, Turgay K, Mechtler K, Charpentier E, Clausen T. McsB Is a protein arginine kinase that phosphorylates and inhibits the heat-shock regulator CtsR [J]. *Science*, 2009, 324:1323–1327.
[38] Wilson ME, Consigli RA. Characterization of a protein kinase activity associated with purified capsids of the granulosis virus infecting Plodia interpunctella [J]. *Virology*, 1985, 143:516–525.
[39] Sikorska M, Whitfield JF. Isolation and purification of a new 105 kDa protein kinase from rat liver nuclei [J]. *Biochim Biophys Acta*, 1982, 703:171–179.
[40] Levy-Favatier F, Delpech M, Kruh J. Characterization of an arginine-specific protein kinase tightly bound to rat liver DNA [J]. *European J Biochem FEBS*, 1987, 166:617–621.
[41] Zetterqvist O, Engstrom L. Isolation of N-e- 32P phosphoryl-lysine from rat-liver cell sap after incubation with 32P adenosine triphosphate [J]. *Biochim. Biophys. Acta,* 1967, 141:523–532.
[42] Chen CC, Smith DL, Bruegger BB, Halpern RM, Smith RA. Occurrence and distribution of acid-labile histone phosphates in regenerating rat liver [J]. *Biochemistry*, 1974, 13:3785–3789.
[43] Chen CC, Bruegger BB, Kern CW, Lin YC, Halpern RM, Smith RA. Phosphorylation of nuclear proteins in rat regenerating liver [J]. *Biochemistry*, 1977, 16:4852–4855.
[44] Fazekas S, Hutas I, Ovary I, Szekessy-Heymann V. Presence of covalently bound energy-rich phosphates in human tracheal smooth muscle myosin [J]. *Acta Physiol Hung*, 1987, 70:385–395.
[45] Zaia DAM, Zaia CTBV, De Santana H. Which amino acids should be used in prebiotic chemistry studies? [J]. *Origins Life Evol B*, 2008, 38:469–488.
[46] Miller SL. A production of amino acids under possible primitive earth conditions [J]. *Science*, 1953, 117:528–529.
[47] Westheimer FH. Why nature chose phosphates [J]. *Science*, 1987, 235:1173–1178.
[48] Bowler MW, Cliff MJ, Waltho JP, Blackburn GM. Why did Nature select phosphate for its dominant roles in biology? [J]. *New J Chem*, 2010, 34:784–794.
[49] Schwartz AW. Phosphorus in prebiotic chemistry [J]. *Phil Trans R Soc B*, 2006, 361:1743–1749.
[50] Todd AR: *Where there's Life there's Phosphorus* [M]. Tokyo, Japan: Sci. Soc. Press; 1981.
[51] Duve DC: *Blueprint for a Cell* [M]. Burlington: Carolina Biological Supply Company, 1991.
[52] Gulick A. Phosphorus as a factor in the origin of life [J]. *American Scientist*, 1955, 43.
[53] Keefe AD, Miller SL. Are polyphosphates or phosphate esters prebiotic reagents? [J]. *J Mol Evol*, 1995, 41:693–702.
[54] Pasek MA. Rethinking early Earth phosphorus geochemistry [J]. *Proc Natl Acad Sci USA*, 2008, 105:853–858.
[55] Adcock CT, Hausrath EM, Forster PM. Readily available phosphate from minerals in early aqueous environments on Mars [J]. *Nature Geoscience*, 2013, 6:824–827.
[56] Yamagata Y, Watanabe H, Saitoh M, Namba T. Volcanic production of polyphosphates and its relevance to prebiotic evolution [J]. *Nature*, 1991, 352:516–519.
[57] Krishnamurthy R, Arrhenius G, Eschenmoser A. Formation of glycolaldehyde phosphate from glycolaldehyde in aqueous solution [J]. *Orig Life Evol Biosph*, 1999, 29:333–354.
[58] Mullen LB, Sutherland JD. Formation of potentially prebiotic amphiphiles by reaction of beta-hydroxy-n-alkylamines with cyclotriphosphate [J]. *Angew Chem Int Ed*, 2007, 46:4166–4168.
[59] Ni F, Sun S, Huang C, Zhao Y. N-phosphorylation of amino acids by trimetaphosphate in aqueous solution-learning from prebiotic synthesis [J]. *Green Chem*, 2009, 11:569–573.
[60] Rabinowitz J, Flores J, Kresbach R, Rogers G. Peptide formation in the presence of linear or cyclic polyphosphates [J]. *Nature*, 1969, 224:795–796.

[61] Ni F, Gao X, Zhao Z-X, Huang C, Zhao Y-F. On the Electrophilicity of Cyclic Acylphosphoramidates (CAPAs) Postulated as Intermediates [J]. *Eur J Org Chem*, 2009:3026–3035.
[62] Gao X, Ni F, Bao J, Liu Y, Zhang Z, Xu P, Zhao Y. Formation of cyclic acylphosphoramidates in mass spectra of N-monoalkyloxyphosphoryl amino acids using electrospray ionization tandem mass spectrometry [J]. *J Mass Spectrom*, 2010, 45:779–787.
[63] Li YM, Yin YW, Zhao YF. Phosphoryl group participation leads to peptide formation from n-phosphorylamino acids [J]. *Int J Pept Protein Res*, 1992, 39:375–381.
[64] Ma XB, Zhao YF. Phosphoryl group participation in the reactions of n-phosphoryldipeptide acids [J]. *Phosphorus Sulfur Silicon Relat Elem*, 1992, 66:107–114.
[65] Li YM, Zhao YF. The bioorganic chemical-reactions of n-phosphoamino acids without side-chain functional-group participated by phosphoryl group [J]. *Phosphorus Sulfur Silicon Relat Elem*, 1993, 78:15–21.
[66] Xue CB, Yin YW, Zhao YF. Studies on phosphoserine and phosphothreonine derivatives – n-diisopropyloxyphosphoryl-serine and n-diisopropyloxyphosphoryl-threonine in alcoholic media [J]. *Tetrahedron Lett*, 1988, 29:1145–1148.
[67] Yang HJ, Jian L, Zhao YF. N- S phosphoryl migration in phosphoryl glutathione [J]. *Int J Pept Protein Res*, 1993, 42:39–43.
[68] Zhou WH, Ju Y, Zhao YF, Wang QG, Luo GA. Simultaneous formation of peptides and nucleotides from N-phosphothreonine [J]. *Orig Life Evol Biosph*, 1996, 26:547–560.
[69] Zhao YF, Cao PS. Phosphoryl amino-acids – common origin for nucleic-acids and protein [J]. *J Biol Phys*, 1994, 20:283–287.
[70] Fu H, Li ZL, Zhao YF, Tu GZ. Oligomerization of N,O-bis(trimethylsilyl)-alpha-amino acids into peptides mediated by o-phenylene phosphorochloridate [J]. *J Am Chem Soc*, 1999, 121:291–295.
[71] Hou J-B, Zhang H, Guo J-N, Liu Y, Xu P-X, Zhao Y-F, Blackburn GM. Chirality at phosphorus in pentacoordinate spirophosphoranes: stereochemistry by X-ray structure and spectroscopic analysis [J]. *Org Biomol Chem*, 2009, 7:3020–3023.
[72] Zeng ZP HJ, Zhang H & Zhao YF. New feature in the chiral bisamino acyl pentacoordinate spirophosphoranes [J]. *Sci China Chem*, 2010, 7:878–887.
[73] Gao X, Liu Y, Xu PX, Cai YM, Zhao YF. Alpha-Amino acid behaves differently from beta- or gamma-amino acids as treated by trimetaphosphate [J]. *Amino Acids*, 2008, 34:47–53.
[74] Chen ZZ, Tan B, Li YM, Zhao YF, Tong YF, Wang JF. Activity difference between alpha-COOH and beta-COOH in N-phosphorylaspartic acids [J]. *J Org Chem*, 2003, 68:4052–4058.
[75] Chen ZZ, Li YM, Ma J, Tan B, Inagaki S, Zhao YF. Activities of alpha-COOH vs gamma-COOH in N-phosphoryl amino acids: A theoretical study [J]. *J Phys Chem A*, 2002, 106:11565–11569.
[76] Lightstone FC, Bruice TC. Ground state conformations and entropic and enthalpic factors in the efficiency of intramolecular and enzymatic reactions .1. Cyclic anhydride formation by substituted glutarates, succinate, and 3,6-endoxo-Delta(4)-tetrahydrophthalate monophenyl esters [J]. *J Am Chem Soc*, 1996, 118:2595–2605.
[77] Xiao X, Zhang Y. Life in extreme environments: Approaches to study life-environment co-evolutionary strategies [J]. *Sci China Earth Sci*, 2014, 57:869–877.
[78] Liu Y. Seryl-histidine dipeptide: An original molecule evolutionary model of modern protease [J]. *Sci China Chem*, 2011, 41:579.
[79] Han DX, Wang HY, Ji ZL, Hu AF, Zhao YF. Amino acid homochirality may be linked to the origin of phosphate-based life [J]. *J Mol Evol*, 2010, 70:572–582.
[80] Han D, Chen W, Han B, Zhao Y. A new theoretical model for the origin of amino acid homochirality [J]. *Sci China Ser C-Life Sci*, 2007, 50:580–586.

4 Nucleoside-protein coevolution and the origin of genetic code

The origin of genetic code is a fundamental question in biological science. As the smallest reaction mode, *N*-phosphoryl amino acids exhibit good catalytic activity for the formation of both nucleotides and the dipeptide. The diverse reactivity of *N*-phosphoryl amino acids enlightened us to study the interrelationship between phosphorus, amino acids and nucleoside, and the origin of genetic code.

4.1 Major phase in the origin of life

The ancient Earth provides basic material conditions (primordial atmosphere without oxygen), energetic conditions (sustaining radiation, volcanic eruption and lightning) and suitable environment (primordial ocean) for the origin of primal lives [1]. The nonliving matters in the primordial Earth evolve into primordial lives via a slow but complicated processes including:

(1) The transformation of inorganic compounds to small organic compounds
The Miller-Urey experiment proved the transformation of inorganic compounds to organic compounds under the conditions of ancient Earth . The artificial synthesis of adenine reported by Oró et al. in 1960 [2] and the synthesis of ribose reported by Butlerov in 1861 [3] also evidenced the first process of the transformation. Hydrocyanic acid, obtained in the lightning, is an important precursor in the formation of nucleotides and related organic compounds.

(2) From small organic compounds to biomacromolecules
Nucleosides and proteins are two important biomacromolecules in the evolution of lives. There are two hypothesis in origin of life studies: (1) one is the continental origin hypothesis [4] and (2) another is the oceanic origin hypothesis [5]. The former proposed that high temperature conditions near the volcano offered a suitable environment for the polymerization of small organic molecules to form macromolecules. Then rain drop erosion brought these primordial macromolecules into the ancient ocean. Fox et al. and Harada et al. [6] reported the transformation of amino acids into polymer under high temperature. These experiments indicated that high temperature may play a significant role in the polymerization of small molecules. Harada et al. also held the view that this polymeric process would took place in the ancient ocean, where small organic molecules attached on the activated surface of clay polymerized in the presence of condensing reagents. Akabori et al. reported that the proteins obtained in a submarine volcano can be synthesized by heating the mixture of amino acids, water and clay. [7] Katchalsky et al. reported the montmorillonite-catalyzed polymerization of amino acids [8].

https://doi.org/10.1515/9783110562552-004

(3) From biomacromolecules to multi-molecule system
The full aggregation and connection of biomacromolecules led to the formation of original system and original lives. The model study indicated that aggregation was a possible pathway for the origin of primordial membrane [9]. Fox et al. reported the formation of microsphere by heating amino acids at 65°C in the presence of phosphoric acid [10]. However, this multi-molecule model was regarded as an insignificant model in the ancient Earth study due to the condition limit.

(4) From multi-molecule system to primal lives
The "aggregate" or "microsphere" cannot be considered as definite lives. There are two hypothesis for the origin of primal life: (1) protein was generated in the first step of the life evolution [11] and (2) nucleosides were generated first and became the fundament of all biomolecules in the chemical evolution that followed [12]. Nucleosides and proteins showed remarkable interaction with each other, and they play irreplaceable roles for lives [13]. The chemical evolution processes of the ancient lives were concluded in Figure 4.1.

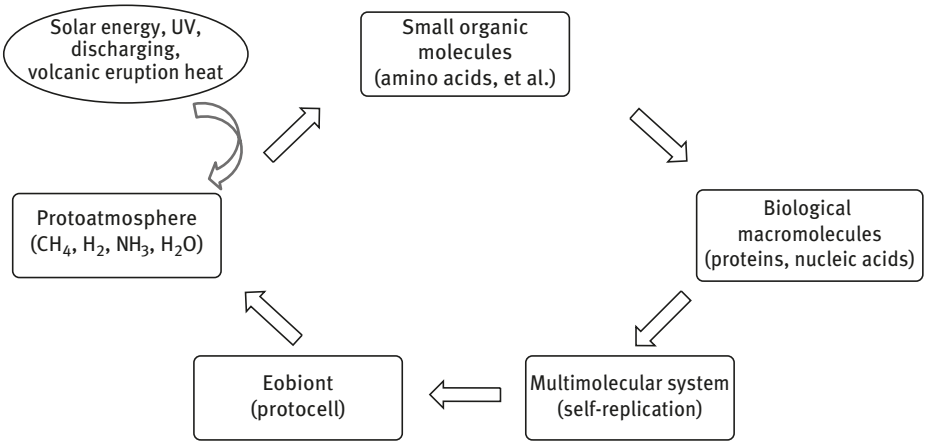


Figure 4.1: Chemical evolution in the origin of life.

Nucleotides and proteins are two types of biopolymer that play important roles in the origin of life. There are two hypotheses concerning the origin of nucleosides and proteins. In the "RNA world" hypothesis, nucleosides are regarded as the information carriers and are able to play the role of proteins, including the direction of protein synthesis [14]. The "protein hypothesis" considered that amino acids can be easily produced under conditions of primitive Earth . Peptides and proteins, obtained from the condensation of amino acids, are able to catalyze nearly all biochemical reactions [15]. The "protein hypothesis" insists that proteins are the basis of lives, including the protein-catalyzed nucleoside synthesis. Therefore, debates between these two

hypotheses were continual for years and endless like the "chicken egg problem". Despite the primary experimental basis, both the two hypotheses are limited in the explanation of its logical integrity. For "protein hypothesis," the self-reproduction of protein remains mysterious, whereas the low catalytic efficiency and limited substrate scope in RNA-enzyme-catalyzed reaction hindered the completeness of "RNA world" hypothesis. In addition, the formation of RNA under prebiotic conditions is still unsolved.

Hence, the key scientific problem in the origin of life study is the origin of genetic code.

4.2 Origin of genetic code

Through careful comparisons and statistical assessment on rRNA sequences, Eigen proposed that the genetic code appeared 3.6 billion years ago. Continuous evolution of genetic code provided the modern genetic code table with relative fixation. In 1958, Crick proposed the central dogma for the genetic expression (Figure 4.2) [16]. The hereditary information is transmitted from DNA to RNA, then from RNA to proteins and in some case from RNA to DNA. The manner of hereditary information transmission is universal in all kind of cells from higher organisms such as human being to lowly life forms such as bacteria. The transmission of hereditary information from DNA to RNA was defined as transcription. The transcription product (initial RNA) then processed into the mature RNA, which is the so-called message RNA (mRNA). The hereditary information containing the mRNA is further used in the directive protein synthesis – the translation process.

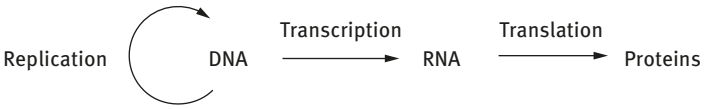


Figure 4.2: The central dogma.

The amino acid sequence was synthesized according to the nucleoside sequence of mRNA. The translation process started from the initiation codon AUG, every three nucleoside in sequence consisted a codon, corresponding to an amino acid. The amino acid sequence of the protein was directed by the genetic codon of mRNA. Despite some lower organisms, most of the organisms share a universal genetic code table.

mRNA is constituted by four different nucleoside, including adenosine (A), guanine (G), cytidine (C) and uridine (U). Here comes the question, why a genetic

codon containing three nucleosides but not one or two? What is the relation between the amino acid and three corresponding nucleosides? In 1961, Matthaei and Nirenberg from NIH deciphered the first genetic codon through a lot of experiments. Man-made RNA – poly U was added into a cell-free extracting solution of *Escherichia coli*, a stable system containing necessary amino acids, DNA, RNA, ribosome and other organelles. PolyPhe was obtained through the resulting protein extracting solution [17]. Thus, the polynucleoside was proved to be directed in the synthesis of protein from free amino acids and the corresponding amino acid codon UUU was proved to be phenylalanine. The genetic codon of proline (CCC), lysine (AAA) was then discovered by the same research group. Following this landmark discovery of Matthaei and Nirenberg, modified methods were employed in the discovery of the rest of the genetic codon (Table 4.1) [18]. In this genetic code table, except the three ending codon, the rest 61 codons were more than the number amino acids in proteins. However, if one amino acid is related to only one genetic codon, there should be some genetic codons that have no corresponding amino acid. On the other hand, if one amino acid is related to two genetic codons, there is no enough genetic codon for each of the 20 amino acids. Therefore, experimental method is the only approach toward the discovery of genetic codon table.

Table 4.1: Tabular form of the genetic codon of higher organisms.

	U	**C**	**A**	**G**	
U	Phe	Ser	Tyr	Cys	**U**
	Phe	Ser	Tyr	Cys	**C**
	Phe	Ser	STOP	STOP	**A**
	Phe	Ser	STOP	Trp	**G**
C	Leu	Pro	His	Arg	**U**
	Leu	Pro	His	Arg	**C**
	Leu	Pro	Gla	Arg	**A**
	Leu	Pro	Gla	Arg	**G**
A	Ile	Thr	Asn	Ser	**U**
	Ile	Thr	Asn	Ser	**C**
	Ile	Thr	Lys	Arg	**A**
	Met	Thr	Lys	Arg	**G**
G	Val	Ala	Asp	Gly	**U**
	Val	Ala	Asp	Gly	**C**
	Val	Ala	Glu	Gly	**A**
	Val	Ala	Glu	Gly	**G**

Despite tryptophan and methionine, other amino acids shown in Table 4.1 were coding by more than one genetic codon. Some amino acids such as leucine, serine

or arginine even have six corresponding genetic codon. As shown in Table 4.1, genetic codons shared the same first two nucleosides for most of the amino acids. For example, glycine corresponded to the following four genetic codons: (1) GGA, (2) GGG, (3) GGC and (4) GGU. In other words, for the genetic codon triplet, the first two nucleosides are much more important than the last one; the coding of amino acids is more related to the first two nucleosides, especially the middle one. Scientists proposed different theories to explain the genetic codon. For example, "3 read 1" theory proposed by Crick [19], "3 read 2" theory proposed by Lagerkvist [20] and "3 read 3" theory proposed by Woese [21].

4.3 The evolution of genetic code

What is the relationship between the genetic code triplet amino acids? The origin of genetic code is one of the most important scientific problems in the world that attracted many attentions. Although the genetic code table was decoded in the 1960s, the origin of genetic code, which was the core issue in the study of origin of life, still remains mysterious. In earlier study of the origin of genetic code, there are two opposite theories, one is the stereochemical theory proposed by Woese in 1966 [22] and another is the ancient frozen theory proposed by Crick in 1968 [23]. The stereochemical theory regards the stereochemical interaction between nucleoside with amino acid as the driving force in the origin of genetic code. Woese also pointed out the importance of hydrophobicity and hydrophilicity in the origin of genetic code. While studying the relationship between amino acids and the corresponding genetic code, Woese et al. found that the importance of the three base groups in the triplet is based on their location. Based on the intensity of the interaction between the second base group and the side chain of the amino acid (hydrophobicity and hydrophilicity), the second base plays the principal role in the triplet. Once the second base group changed to another base group, the corresponding amino acid is changed. However, the first base group showed a lower significance than the second one, whereas the last base group showed less influence. When the last nucleoside of the genetic code triplet was changed, the corresponding amino acid would not be changed or just be changed to another amino acid with similar polarity. Illuminated by the hierarchy interactions, Woese proposed that the second base of the triplet shows strong stereochemical interaction with the corresponding amino acid side chain, the resulting triplet-amino acid adducts decided the correspondence.

Crick et al. regarded the relation between genetic code triplet and amino acid as a casual phenomenon, which was fixed during the long-term evolution, and is difficult to be modified. In this hypothesis, Crick proposed that, in the early stage, the genetic codons contained only one base group. During the evolution, the genetic

codons were evolved, doublet genetic codons appeared, and finally the triplet genetic codons. Once the genetic code in organism was changed, it led to serious problem that may have caused extinction. However, Crick regards the genetic codon table as a ruleless table. Thus, this hypothesis cannot explain the hierarchical influence of the genetic codon. The ancient frozen theory was also been challenged by other theories. For example, in 1981, Eigen proposed the "in vitro selection" theory, in which the original genetic codons contained more G, C base groups than A, U base groups. According to this theory, the codes G, C appeared first and made up the earlier genetic code system. Then the code A was joined when the old GC system was replaced by a new GCA system, resulting in the increase of corresponding amino acids. Finally, code U was obtained in the genetic code system and the modern GCAU system formed [24]. In 1989, Orgel proposed the decoding mechanism origin theory [25], in which the adjacent two triplets appeared at the same time with the corresponding adjacent two amino acids during the evolution. In 1988, Christian de Duve published his second genetic code hypothesis [26]. Later in 1999, Knight et al. reported a comprehensive hypothesis on the genetic code evolution [27]. Knight proposed that the stereochemistry and coevolution displayed different roles in different periods.

The genetic code coevolutionary theory proposed by Wong et al. is one of the greatly influenced theory in the genetic code origin study [28]. Following this, Di Giulio [29] reported a modified coevolutionary theory in 2004. Wong et al. also discovered that there is only one base group difference between the biosynthetic amino acids and its amino acid precursors. Thus, Wong et al. proposed that the original simple genetic code was evolved together with the original amino acid biosynthesis. For example, when alanine was transformed to glycine, the genetic code GCA (Ala) was changed into GGA (Gly). The genetic code of precursor amino acid will completely or partly change into the code of the product amino acid, whereas the precursor amino acid Ala gradually transformed into the product amino acid Gly. The coevolution of amino acids with genetic codes was displayed in Figure 4.3.

In 1997, Amimovin et al. [30] verified the coevolution hypothesis via theoretical methods and proved a predicted genetic code table that was randomly generated by computer. Numerical analysis showed that the biosynthesis of original amino acids had a great influence on the formation of genetic code table. However, the reason why there is only one base group difference between the biosynthetic amino acids and its amino acid precursors is still unexplained.

In conclusion, despite some lower organisms, most of the organisms share a universal genetic code table. However, the genetic code evolved through the evolution of lives. Concerning the evolution of genetic code, two hypothesis including gradual evolution and random evolution are proposed to solve this problem. Nevertheless, these two hypotheses remain debatable and need experimental support. The study on the origin of genetic code and evolution still need to be improved; meanwhile, new pathway and technology are necessary for further research on this area.

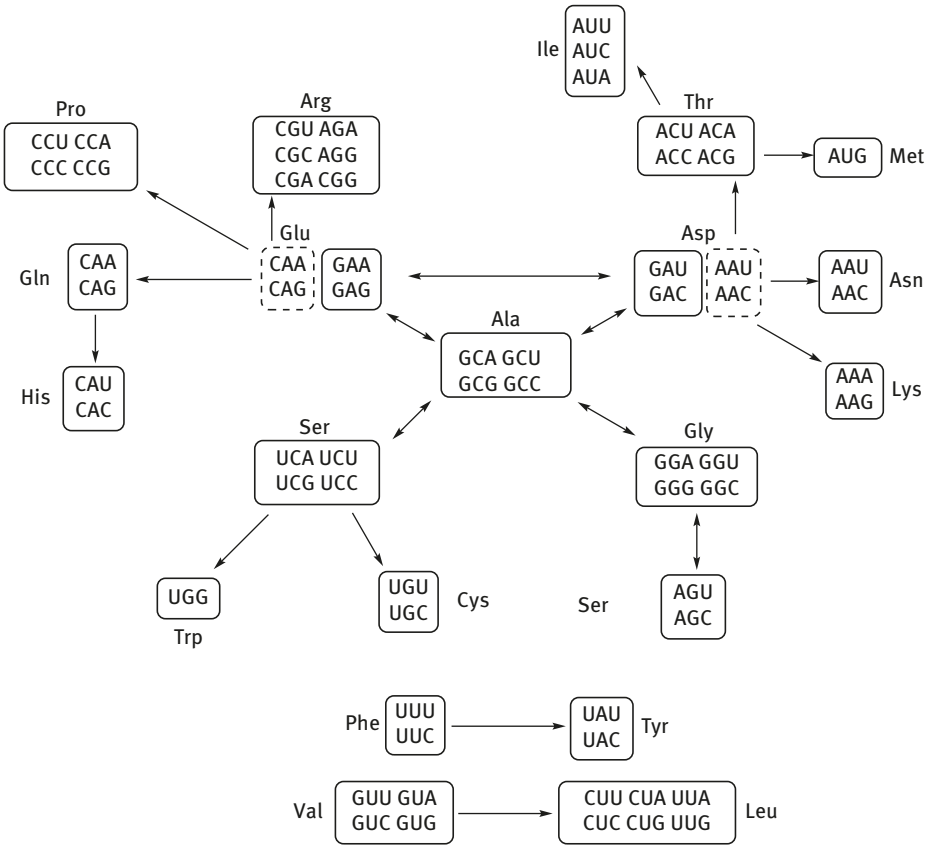


Figure 4.3: Coevolution of amino acids with genetic codes.

4.4 Phosphorus plays an important role in the origin of life

Phosphorus is one of the basic elements for organisms. Despite the low content in organism's body, phosphorus plays an important role in the biochemical reactions. In human body, 1% content of phosphorus is the major component of bones and teeth. Phosphonates are also the linkage of DNA and RNA framework, in which the content of phosphorus is 9%. Meanwhile, the phospholipid bilayer is the major component of the cell membrane, which participates in the regulation of membrane permeability. ATP acts as the energy transformation form in living cells, whereas phosphorylation/dephosphorylation of proteins is the signal transduction mode of lives.

Hence, phosphorus should not be ignored in the origin of life study. It is proved that, under the conditions of the ancient Earth , phosphorus could have promoted the formation of amino acids, nucleosides/nucleotides, basic groups and peptides. As pointed out by Prof. Alexander R. Todd: "If life exists anywhere else in the universe it

will do so only on a planet on which phosphorus is readily available" [31]. Phosphorus is the basic condition for lives. In 1987, Prof. Westheimer from Harvard University demonstrated the problem "why nature choose phosphonates" [32]. Prof. Krebs and Prof. Fischer also reported that the activity of proteins is regulated by reversible phosphorylation [33]. Furthermore, phosphorus can also act as a judgment standard to predict the existence of life outside the Earth. With the development of space science, double atom interstellar molecules PN and PC were obtained [34]. In 2013, Adcock et al. discovered that the content of phosphorus in Mars is 10 times more than that in the Earth [35]. Hence, phosphorus plays an important role in the origin of life research.

In 1953, Miller et al. reported his famous experiment of discharging under ancient primitive atmospheric conditions. After carefully analyzing the reaction products, several small biomolecules such as amino acids were obtained. Following this, other scientists modified the atmospheric composition and the excited conditions (radiation, UV, heat, shock wave, etc.) and reported several unusual discoveries [36].

In 1976, Ridgway et al. first reported the discovery of Phosphane (PH_3) in the atmosphere of Jupiter [37]. PH_3 is proposed to be one of the components of ancient atmosphere and maybe the phosphorus source for lives. Prof. Wang first introduced PH_3 into the simulated ancient atmosphere and the discharging reaction shows that, for the prebiotic synthesis of amino acids, the experiments in the presence of PH_3 resulted in the formation of 14 protein amino acids, whereas the one without PH_3 only resulted in the formation of 6 amino acids [38]. These results indicated that phosphorus is able to promote the prebiotic synthesis of amino acids. Moreover, phosphorus is supposed to be the catalyst for evolution of ancient Earth. Our research group have focused on this area for years and have explained this question in both experimental and theoretical methods [39]. Only α-amino acids are able to form pentacoordinated phosphorus intermediates, whereas β-amino acids cannot proceed the same reaction (Figure 4.4). In addition, a series of experiments was reported that indicated that phosphorus displayed crucial influence on the prebiotic synthesis of small biomolecules (amino acids, nucleosides and basic groups) and biomacromolecules (peptides and nucleotides).

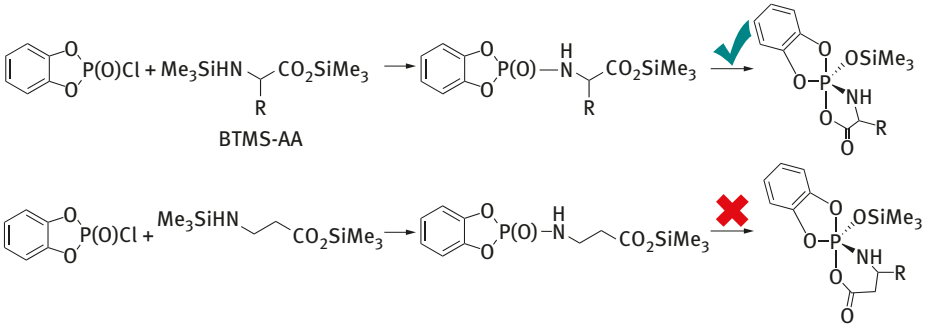


Figure 4.4: Pentacoordinated phosphorus intermediates.

4.5 *N*-phosphoryl amino acids: model for the study of nucleotide-protein coevolution

Concerning the hypotheses on the study of origin of life, indeed, there is no suitable model that can well explain the problems of the origin of life. Hence, our research group employs *N*-phosphoryl amino acids as the model compound to study the interaction (Figure 4.5) [40].

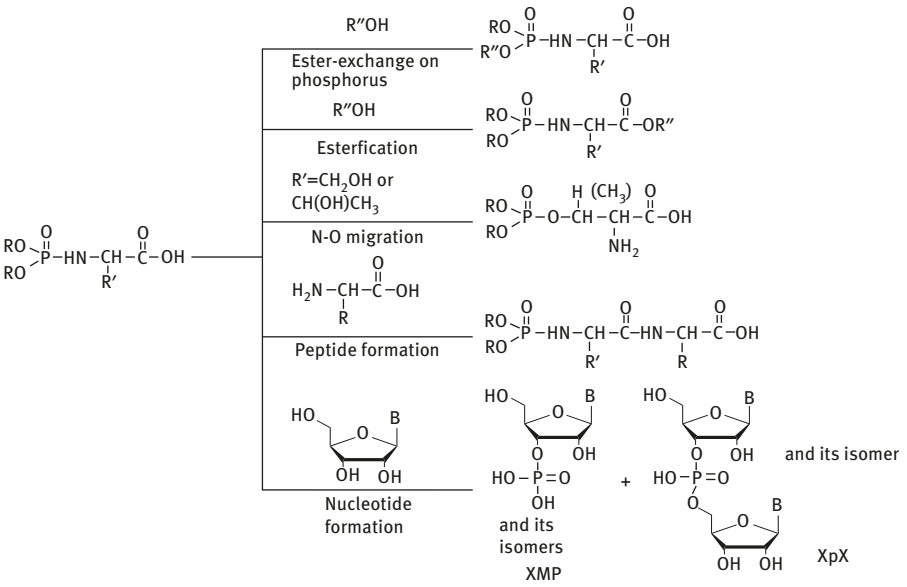


Figure 4.5: Chemical reactivity of *N*-phosphoryl amino acids.

N-phosphoryl amino acids show higher reactivity than free amino acids; they are able to react with alcohols at room temperature and resulted in the formation of ester condensation products. If *N*-phosphoryl amino acids is added into a solution of alcohol and kept in room temperature for a period of time, the hydroxyl group of the alcohol was obtained to react with the phosphorus center and the ester exchange and ester condensation products were detected. Substrate amino acids with hydroxyl groups in the side chain in the reaction lead to an intermolecular phosphoryl transfer (N→O) product *O*-phosphoryl amino acids. In addition, if *N*-phosphoryl amino acids are exposed under mild conditions, they are autotransferred into peptides via an intermediate with pentacoordinated phosphorus center. Moreover, when the reaction took place in the presence of nucleosides, nucleotides and corresponding derivatives were obtained. The reactivity of *N*-phosphoryl amino acids is varied; *N*-phosphoryl

amino acids can act as the endonuclease for DNA. Moreover, a substituted long chain alkoxyl group on the phosphorus center will significantly change the physical behaviors of *N*-phosphoryl amino acids and led to the formation of cell membrane analogues. Compared with the phospholipid membrane, Raman spectra analysis of this membrane analogue showed limited mobility that could act as a model system for the study of membrane.

Based on this smallest chemical model system employing *N*-phosphoryl amino acid as an active center, we reported the bifunctional reactivity of *N*-phosphoryl amino acids toward nucleosides to form nucleotides and toward free amino acids to form dipeptides, respectively. Hence, a hypothesis named "nucleoside-protein co-evolution" was reported by our group [41].

The reaction of *N*-phosphoryl amino acids toward nucleoside led to the formation of complexes containing 2′,3′-dioxaphospholane-nucleoside and amino acid. The complex can react with another molecule of nucleoside and produce a dinucleotide-phosphoryl amino acid complex. The resulting complex could further react with another amino acid and produce the corresponding dinucleotide-phosphoryl dipeptide complex. After several cycles, the oligonucleotides and phosphoryl peptides were detected in the reaction mixture (Figure 4.6). The oligomerization reaction can occur in both organic and aqueous solutions. *N*-phosphoryl amino acids display selectivity toward different nucleoside in the phosphorylation. A mechanism was proposed to explain the chemical selectivity: the weak interaction between the side chain of amino acids and the basic group of nucleoside (hydrogen bond, π-π interaction, Van der Waals' force and hydrophobicity) will cause different selectivity on the substrates. In summary, phosphorus is the regulation center in the origin and evolution of amino acids and nucleosides. Meanwhile, this chemical model opens a new avenue in the origin of genetic code study.

4.6 The relationship of dipeptide yields with nucleoside in the phosphorus-assisted condensation of amino acids

As one of the important components in organism, peptides act as the chemical messenger that carries information and enters the cell via membrane receptor. In human body, peptide participated in many important physiological processes, such as blood pressure regulation, breathing, digestion, reproduction, metabolism, immunity and algesia conduction. The peptide research always attracts much attention in the biological science.

Phosphorus reagents have wide applications in the peptide formation [42]. Amido bonds are formed via the intermediate attracted by the amino group and the activation of carbonyl group. Fukuji et al. [43] reported poly amido products by treating amino acids with $POCl_3$. We have studied the self-assembled reactivity of 20 natural

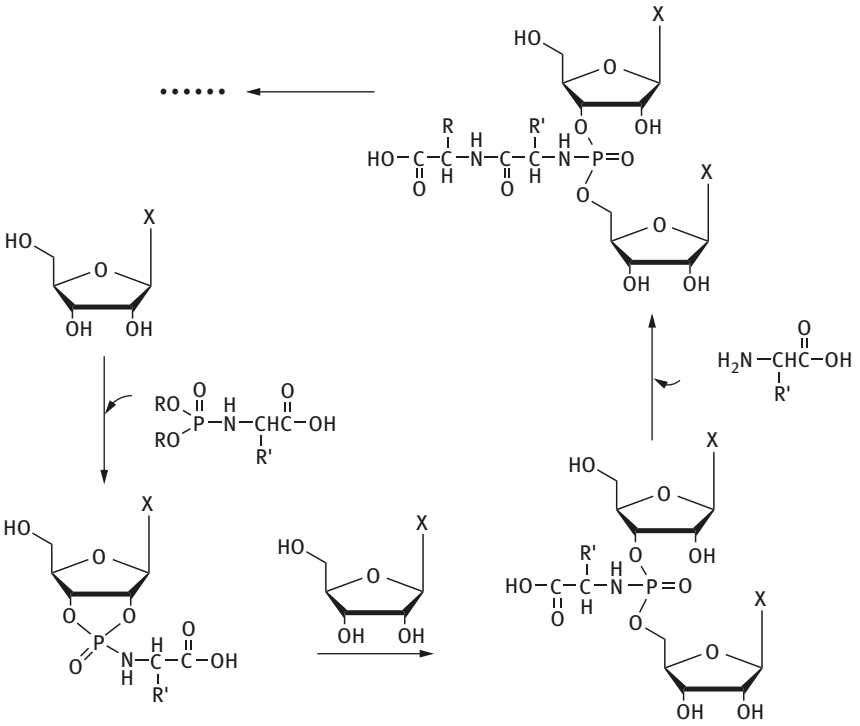


Figure 4.6: Formaiton of *N*-phosphoryl peptides and nucleotides.

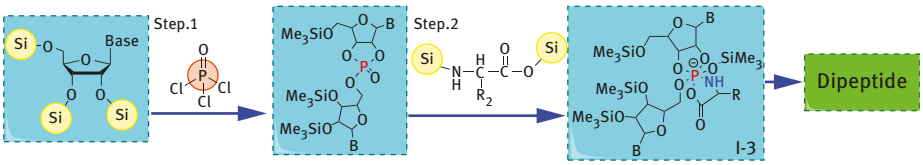


Figure 4.7: Chemical model for the study of the relationship between dipeptide yields and nucleoside in the phosphorus-assisted condensation of amino Acids.

amino acids in the presence of $POCl_3$. Further, inorganic phosphorus compounds also proved to be efficient reagents for the peptide formation from free amino acids [44]. As described previously, the middle nucleoside of the genetic code triplet has a very strong relationship with the corresponding amino acid. As possible condensation reagents for peptide formation under the prebiotic conditions, inorganic phosphorus exhibited important meaning in the research of the origin of life. Hence, the study of genetic code origin through comparison of dipeptide yields with nucleoside would be a convenient and efficient research method (Figure 4.7).

References

[1] McClung C E. The unity of life[J]. *Science* 1926, *64*: 561–569.
[2] Oró J, Kimbal A P. Synthesis of purines under possible primitive earth conditions. I. Adenine from hydrogen cyanide[J]. *Arch. Biochem. Biophys*. 1961, *94*: 217–227.
[3] Butlerow A. Bildung einer zuckerartigen Substanz durch Synthese[J]. *Justus Liebigs Annalen der Chemie* 1861, *120*: 295–298.
[4] Gilvarry J J. Thermal history of the Moon[J]. *Nature* 1970, *225*: 623–625.
[5] Melton F A. Linear and denritic sink-hole patterns in Southeastern New Mexico[J]. *Science* 1934, *80*: 123–124.
[6] Vegotsky A, Harada K, Fox S W. Characterization of polyaspartic acid and some related compounds[J]. *J. Am. Chem. Soc*. 1958, *80*: 3361–3363.
[7] Akabori S, Mizushima S. *Protein Chemistry*. 1st edn. [M]. Kyoritsu Shuppan: Tokyo, 1954.
[8] Paecht-Horowitz M. The mechanism of clay catalyzed polymerization of amino acid adenylates[J]. *Biosystems* 1977, *9*: 93–98.
[9] Bungenberg de Jong H. G. In *Colloid Science*. Kruyt H R. Ed. [M]. Elsevier: London, 1949, 2.
[10] Fox S W. *The Origins of Prebiological Systems and of Their Molecular Matrices*[M]. Academic Press: New York, 1965.
[11] Oparin A I. *The Origin of Life*[M]. Weidenfeld and Nicolson: London, 1924, 199–234.
[12] Haldane J B S. *The Origin of Life*[M]. Rationalist Annual, 1929.
[13] Lehninger A L. *Biochemistry*[M]. Worth: New York, 1975.
[14] Kruger K, Grabowski P J, Zaug A J, Sands J, Gottschling D E, Cech T R. Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena [J]. *Cell* 1982, 31: 147–157.
[15] Fox S W, Dose K. *Molecular Evolution and the Origin of Life*[M]. New York: Dekker, M., Inc. 1977, 370.
[16] Crick F H C. Central dogma of molecular biology[J]. *Nature* 1970, *227*: 561–563.
[17] Nirenberg M W, Matthaei H. The dependence of cell-free protein synthesis in Escherichia coli upon naturally occurring or synthetic polyribonucleotides[J]. *Proc. Natl. Acad. Sci.* 1961, *47*: 1588–1602.
[18] Crick F H C, Barnett L, Brenner S, Watts-Tobin R J. General nature of the genetic code for proteins [J]. *Nature* 1961, *192*: 1227–1232.
[19] Crick F H C. Codon-anticodon pairing: The wobble hypothesis[J]. *J. Mol. Biol.* 1966, *19*: 548–555.
[20] Lagerkvist U. "Two out of three": An alternative method for codon reading[J]. *Proc. Natl. Acad. Sci.* 1978, *75*: 1759–1762.
[21] Woese C R, Dugre D H, Saxinger W C, Dugre S A. The molecular basis for the genetic code[J]. *Proc. Natl. Acad. Sci.* 1966, *55*: 966–974.
[22] Woese C R, Dugre D H, Dugre S A, Kondo M, Saxinger W C. On the fundamental nature and evolution of the genetic code[J]. *Cold Spring Harbor Symp. Quant. Biol.* 1966, *31*: 723–736.
[23] Crick F H C. Origin of the genetic code[J]. *J. Mol. Biol.* 1968, *38*: 367–379.
[24] Hartman H. Speculations on the evolution of the genetic code[J]. *Origins Life* 1975, *6*: 423–427.
[25] Orgel L E. The origin of polynucleotide-directed protein synthesis[J]. *J. Mol. Evol.* 1989, *29*: 465–474.
[26] de Duve C. Transfer RNAs: The second genetic code[J]. *Nature* 1988, *333*: 117–118.
[27] Knight R D, Freeland S J, Landweber L F. Selection, history and chemistry: The three faces of the genetic code[J]. *Trends Biochem. Sci.* 1999, *24*: 241–247.
[28] Wong J T. Coevolution theory of the genetic code[J]. *Proc. Natl. Acad. Sci.* 1975, *72*: 1909–1972.
[29] Di Giulio M. The coevolution theory of the origin of the genetic code[J]. *Phys. Life Rev.* 2004, *1*: 128–137.

[30] Amimovin R. An analysis of the metabolic theory of the origin of the genetic code[J]. *J. Mol. Evol*. 1997, *44*: 473–476.
[31] Todd A R. In "Where there's life there's phosphorus"[M], Ed. Kageyama, M, Nakamura, K, Oshima, T. Japan Sci. Soc. Press: Tokyo, Japan, 1981, 275.
[32] Westheimer F H. Why nature chose phosphates[J]. *Science*, 1987, *235*: 1173–1178.
[33] Krebs E G, Fischer E H. Phosphorylase activity of skeletal-muscle extracts[J]. *J. Biol. Chem*. 1955, *216*: 113–120.
[34] Weckwerth G, Schidlowski M. Phosphorus as a potential guide in the search for extinct life on Mars[J]. *Adv. Space. Res*. 1995, *15*: 184–190.
[35] Adcock C T, Hausrath E M, Forster P M. Readily available phosphate from minerals in early aqueous environments on Mars[J]. *Nature Geosci.* 2013, 6: 824–827.
[36] Crossenbacher K A, Knight C A. *The Origins of Prebiological Systems*[M]. New York: Academic Press, 1965.
[37] Ridgway S T, Wallace L, Smith G R. The 800–1200 inverse centimeter absorption spectrum of Jupiter[J]. *Astrophys. J.* 1976, *207*: 1002–1006.
[38] Kobayashi K, Ponnamperuma C. Trace elements in chemical evolution. II: Synthesis of amino acids under simulated primitive earth conditions in the presence of trace elements[J]. *Orig. Life Evol. Biosph*. 1985, *16*: 57–67.
[39] Fu H, Li Z.-L, Zhao Y.-F, Tu G -Z. Oligomerization of N,O-Bis(trimethylsilyl)-α-amino Acids into Peptides Mediated by o-Phenylene Phosphorochloridate[J]. *J. Am. Chem. Soc.* 1999, *121*: 291–295.
[40] Ma X, Zhao Y. Synthesis and novel properties of N-phosphoryl peptides[J]. *J. Org. Chem*. 1989, *54*: 4005–4008.
[41] Zhao Y -F, Cao P –S. Basic models of chemical evolution of life: The minimum evolving system. In: Chela-Flores J., Raulin F. Eds. *Chemical Evolution: Physics of the Origin and Evolution of Life*. Netherlands: Kluwer Academic Publishers[M], 1996, 279–285.
[42] Sheehan J C, Frank V S. Peptide syntheses using energy-rich phosphorylated amino acid derivatives[J]. *J. Am. Chem. Soc.* 1950, *72*: 1312–1316.
[43] Higashi F, Fujiwara Y-I, Yamada Y. Direct polyesterification promoted by a complex of phosphorus oxychloride with lithium chloride monohydrate[J]. *J. Polym. Sci. A Polym. Chem*. 1986, *24*: 589–594.
[44] Zhou N, Lu K, Liu Y, Chen Y, Tang G, Cao S -X, Qu L -B, Zhao Y -F Electrospray ionization mass spectrometric studies of phosphorus oxychloride directed synthesis of homo-oligopeptide ester libraries[J]. *Rapid Commun. Mass Spectrom.* 2002, *16*: 919–922.

5 The phosphoryl transfer reactions of phosphoryl amino acids

It is a well-known that many enzymatic functions are regulated by protein phosphorylation and dephosphorylation. This regulation is actually related to phosphoryl transfer reaction. The mechanism of phosphoryl transfer reaction is much more complicated than the reaction associated with carbonyl group. The phosphoryl transfer reactions usually involve phosphorus stereochemistry and pentacoordinated phosphorus intermediate or transition state.

5.1 The research progress of phosphoryl transfer reaction

More than 20 years ago, Prof. Zhao's team discovered that *N*-phosphoryl-α-amino acids could induce many bioorganic reactions [1–5], such as peptide formation in water–alcohol medium [6–8]. However, *N*-phosphoryl-β-or γ-amino acids could not produce self-assembly peptides. Further studies showed that the transition state of *N*-phosphoryl-α-amino acid was a pentacoordinated mixed anhydride derived from phosphoric acid and carboxylic acid. The center of pentacoordinated phosphorus was activated by the carboxyl group; therefore, the ester exchange reaction on phosphorus atom could be carried out. When two ester exchange reactions of hydroxyl groups on nucleoside ribose had been completed, nucleosides were converted to nucleotides [9]. In addition, *N*-phosphoryl serine or threonine with hydroxyl group on the side chain could undergo intramolecular phosphoryl transfer reactions to yield *O*-phosphoryl amino acids [10]. Based on the detailed studies on *N*-phosphoryl amino acids, Prof. Zhao's team first proposed the hypothesis of "protein and nucleic acid common-origin theory" [11, 12]. The hypothesis presumed that protein and nucleic acid all originated from *N*-phosphoryl amino acids, which were derived from the reaction of high-energy phosphate compounds on the ancient Earth reacting with amino acids. Because the ancient Earth had volcano activity, pyrophosphate and pyrophosphate compounds were likely accumulated on the surface of the Earth. The energy of these compounds with P–O–P bond was transferred to peptide and nucleotide phosphate ester bonds through the formation of P–N bond with amino acids. Therefore, *N*-phosphoryl amino acids could be used as a small molecular model to study many unsolved mysteries in the origin of life, such as the origin of chirality, the origin of genetic code and so on. Afterward, the "protein and nucleic acid common-origin theory" based on phosphoryl amino acids has attracted much attention in the world.

The studies on the chemical activities of *N*-phosphoryl-α-amino acids from 20 natural amino acids found that they could induce many kinds of bioorganic reactions in water–alcohol medium, such as ester formation, peptide formation, ester exchange

https://doi.org/10.1515/9783110562552-005

on phosphoryl atom and N→O phosphoryl migration. The ester exchange reaction and N→O migration reaction belong to the transfer reaction of phosphoryl group. That means phosphoryl group can be transferred from a residue of one amino acid to that of another amino acid. Further studies on phosphoryl transfer reaction will be helpful to investigate the function of phosphorus in the origin of life and genetic codon.

The phosphoryl transfer reactions often occur in the phosphorylation and dephosphorylation processes of enzymes and proteins. These processes play an important role in biochemistry with the regulation of material exchange, energy exchange and transmission of information in human life. The phosphoryl transfer reaction includes O→O phosphoryl migration reaction, which was also known as ester exchange reaction, and N→O phosphoryl migration reaction. The phosphoryl transfer reaction exists widely not only among biological macromolecules but also between small molecules.

The ester exchange reaction, O→O phosphoryl migration reaction, widely occurs in reactions involving phosphoryl kinase. The phosphoryl transfer process actually involves pentacoordinated phosphorus transition state and chiral phosphoryl group [13, 14]. All of the following phosphoryl transfer processes are stereospecific: the hydrolysis process of RNA catalyzed by bovine pancreatic ribonuclease, the process of β-D-glucose phosphorylation by adenosine triphosphate (ATP) under the action of an enzyme and the hydrolysis process of myosin ATP. On the other hand, in the phosphorylation and dephosphorylation processes of enzymes and proteins, there are many types of intermolecular N→O phosphoryl transfer reactions, and the mechanism is considered through a mixed anhydride intermediate of pentacoordinated phosphorus [15]. In phosphoryl protein, when the nitrogen atom in histidine imidazole group combines with ATP, an adenosine diphosphate molecule is eliminated and a single phosphorylation histidine unit is formed. Then, the phosphoryl group is transferred from the nitrogen atom of histidine unit to the oxygen atom of side chains of aspartic acid through a mixed anhydride intermediate of pentacoordinated phosphorus.

Acetylcholinesterase (AChE) is one of the proteins that is sensitive to phosphorylation. Enzyme is aging by phosphorylation of AChE and results in pathological changes. On the other hand, it can be revived by dephosphorylation process and promote functional recovery of the organism. The active center of AChE is composed of side-chain imidazole, hydroxyl and carboxyl groups of histidine, serine and aspartic acids. The imidazole group and the hydroxyl group are almost in the same active position [16]. The phosphorylation process of AChE occurs mainly through the coordination of imidazole group, whereas phosphorus atom attacks the hydroxyl groups to form phosphorylated enzymes. Another viewpoint is that the phosphoryl group first attacks the imidazole base and then transfers to the hydroxyl group. Therefore, the study of the interaction and mechanism between histidine side chain and phosphoryl group is the key to study the mechanism of AChE phosphorylation. In recent years, the investigation of the catalytic mechanism of AChE showed that the synergistic effect of

histidine and serine side chain is the key to the phosphoryl transfer and physiological function. With the determination of AChE molecular structure, the studies on catalysis and aging mechanism-related AChE will be further carried out. Furthermore, the studies will provide a good model for the study of protein phosphoryl transfer mechanism.

Transcription factors are a class of protein factors that bind to specific DNA sequences and regulate transcription process. Studies have shown that phosphorylation of transcription factors regulates their activity [17]. The transcription factor in the cytoplasm has no activity, since it cannot bind to its target DNA sequence. When the transcription factor itself or some fixed protein in the cytoplasm is phosphorylated, it can reach the nucleus and play its role. The binding activity of nuclear transcription factor with DNA can be regulated by the bidirectional phosphorylation. Therefore, phosphorylation plays an important role in the regulation of biological function of transcription factor.

The phosphoryl transfer reaction also occurs between small organic molecules. Studies on phosphoryl amino acids found that when phosphoryl amino acids are in water–ethanol system, the phosphoryl transfer reaction of ester exchange reaction on phosphorus and N→O phosphoryl transfer reaction occur between small molecules. For the phosphoryl amino acids without hydroxyl group on the side chain, monoester and diester exchange reactions will occur mainly in the alcohol system. On the other hand, for the phosphoryl amino acids with hydroxyl group on the side chain, such as serine and threonine, the products are more complicated. In addition to the monoester and diester exchange products, the intramolecular N→O phosphoryl transfer product may appear.

When we synthesized the seryl–histidine (Ser–His) dipeptide from serine mediated by *O*, *O*-phenylene phosphorochloridate, it was found that *O*,*O*-phenylene phosphoryl group could transfer between different amino acids. The expected product was Ser–His dipeptide, but it was found that there was almost equimolar amount of histidyl–histidine (His–His) dipeptide. In order to further understand the properties and mechanisms of phosphoryl transfer reactions, we carried out a study of the phosphoryl transfer reactions between different amino acids.

5.2 Study on the transfer reaction of Intermolecular *O,O*-Phenylenephosphoryl group

5.2.1 Synthesis and analysis of Ser–His dipeptide

It was found that *N*-phosphoryl-α-amino acids could induce many kinds of bioorganic reactions in water–alcohol medium. These reactions resulted in the pentacoordinated phosphorus amino acid intermediates, which were unstable in the water–alcohol

system and could not be easily detected. In order to confirm the pentacoordinated phosphorus intermediates, an interesting experiment based on silicon chemistry was applied to trap them. The trimethyl-silyl group was found to be a good protective group for amino acids, and the amino acids protected by trimethyl-silyl group could oligomerize into peptides mediated by organic phosphorus reagent. This method was used to synthesize Ser–His dipeptide, the world's smallest peptide at present for DNA cleavage [18, 19].

The synthesis of the Ser–His dipeptide assisted by organic phosphorus reagent **1** is shown in Figure 5.1. The *N*,*O*-bis(trimethylsilyl) serine trimethylsilyl ester **2** was phosphorylated by *O*,*O*-phenylenephosphorochloridate **1**, and then the pentacoordinated phosphorus serine intermediate **4** reacted with *N*,*N*-bis(trimethylsilyl) histidine trimethylsilyl ester **5** to synthesize Ser–His dipeptide.

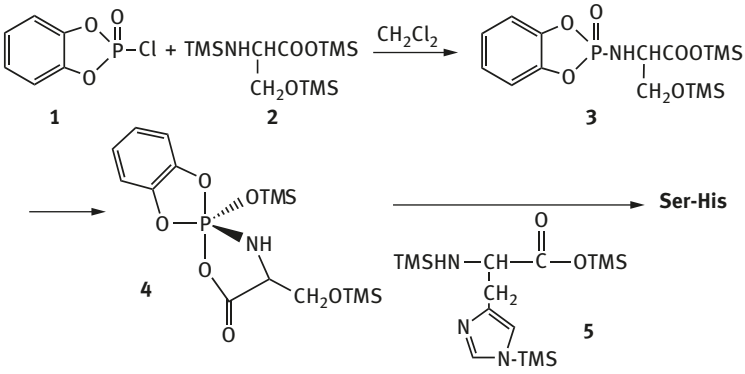


Figure 5.1: Synthetic reaction of Ser–His dipeptide.

The reaction products were analyzed by high-performance liquid chromatography-electrospray ionization-mass spectrometry (HPLC-ESI-MS). The results showed that there was not only Ser–His dipeptide but also His–His dipeptide. The spectra of HPLC-ESI-MS are shown in Figure 5.2.

In all mass spectra of reaction crude products, the peaks at *m*/*z* 156, 243 and 293 were interpreted as $[His + H]^+$, $[Ser\text{-}His + H]^+$ and $[His\text{-}His + H]^+$ ions on the basis of HPLC-ESI-MS/MS, respectively. The peak at *m*/*z* 243 was identified as Ser–His dipeptide, and the peak at *m*/*z* 293 was identified as His–His dipeptide. The contents of dipeptides were calculated by area percent method. The results of analysis showed that there was almost equimolar amount of His–His dipeptide as well as Ser–His dipeptide. Based on HPLC-ESI-MS/MS analysis, we speculated that in the self-assembly peptide reaction of amino acids mediated by organic phosphorus reagent, the phosphoryl transfer process occurred between serine and histidine pentacoordinated phosphorus intermediate.

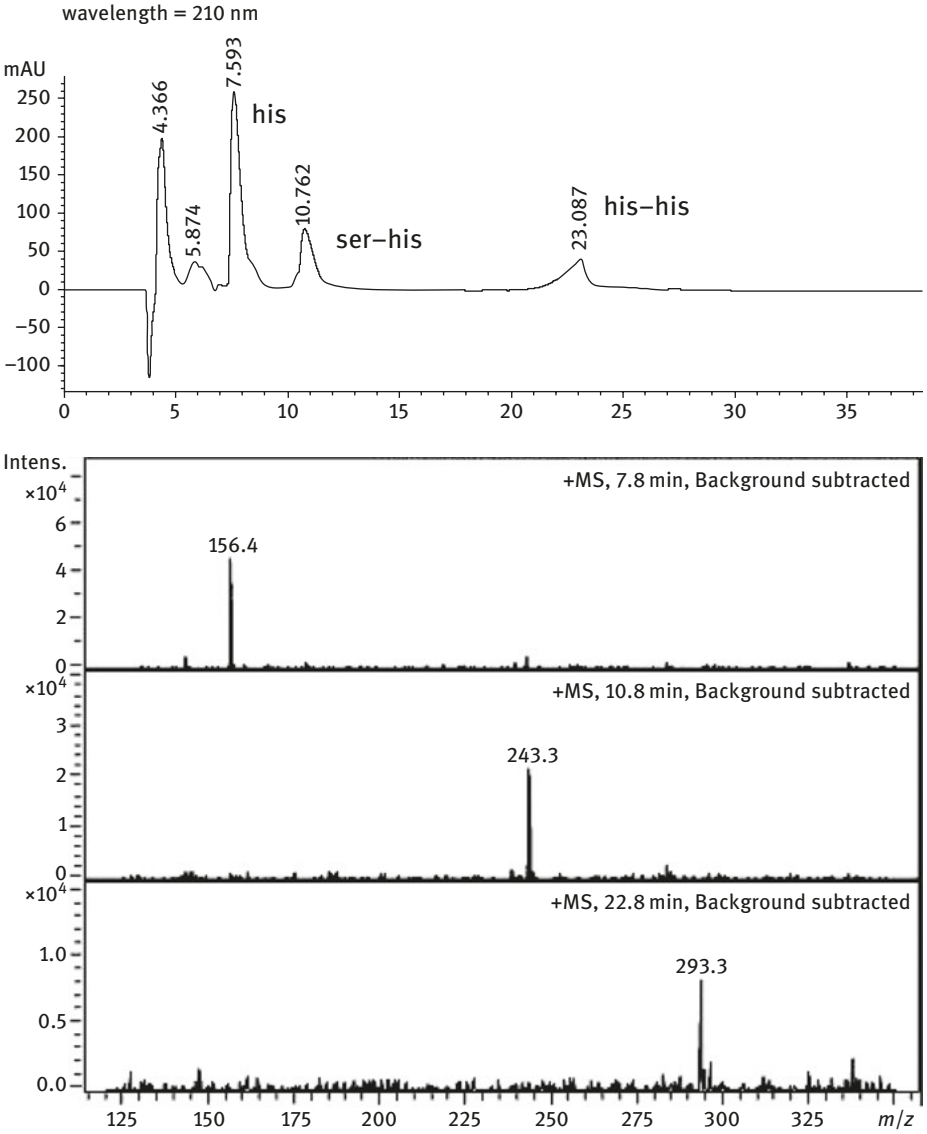


Figure 5.2: HPLC-ESI-MS spectra of reaction crude products.

5.2.2 The investigation of the intermolecular transfer mechanism of *O*,*O*-phenylenephosphoryl group between amino acids

For the self-assembly peptide reaction of *N*,*O*-bis(trimethylsilyl) amino acids mediated by *O*,*O*-phenylene phosphorochloridate, the reaction mechanism comprises three stages: activation of amino acids (the formation of pentacoordinated

phosphorus intermediates), extension of peptide chain and reaction termination [6]. Among the three stages, the first stage, activation of amino acids, is most important. For example, when the *N*,*O*-bis(trimethylsilyl) serine trimethylsilyl ester was activated to pentacoordinated phosphorus intermediate **4**, two likely nucleophilic attack pathways, **a** and **b** pathways, leading to different dipeptides occur as shown in Figure 5.3. Because the basicity of protected imidazole nitrogen was weaker than the protected amino nitrogen, the pathways, **a** and **b**, were the main pathways. After the pentacoordinated phosphorus serine intermediate **4** was formed, the carbonyl carbon of **4** is attacked by the nitrogen atom of *N*,*N*-bis(trimethylsilyl) histidine trimethylsilyl ester **5** to give the phosphoryl Ser–His dipeptide **12**. Then, Ser–His dipeptide **14** was produced after hydrolysis. However, if the phosphorus atom of **4** was attacked by the nitrogen atom of *N*,*N*-bis(trimethylsilyl) histidine trimethylsilyl ester **5**, the pentacoordinated phosphorus histidine intermediate **7** was formed by the substitution of *N*,*N*-bis(trimethylsilyl) histidine trimethylsilyl ester **5** for the *N*,*O*-bis(trimethylsilyl) serine trimethylsilyl ester **2**. That means a competitive *O*,*O*-phenylenephosphoryl group between pentacoordinated phosphorus intermediates **4** and **7** is obtained, resulting in the phosphoryl transfer reaction from serine to histidine. Later, the pentacoordinated phosphorus histidine intermediate **7** was attacked by the nitrogen atom of *N*,*N*-bis(trimethylsilyl) histidine trimethylsilyl ester **5** to give the phosphoryl His–His **8**, and then His–His dipeptide **11** was obtained after hydrolysis. The proposed phosphoryl transfer mechanism for His–His formation is shown in Figure 5.3.

In order to further verify the above phosphoryl transfer mechanism proposed between pentacoordinated phosphorus amino acids, we further explored the mechanism using nuclear magnetic resonance (NMR) tracking technique. The self-assembly reaction of pentacoordinated phosphorus serine **4** was traced by ^{31}P-NMR. It was found that if the reaction temperature was high, there were signals at δ-34.7 and δ-35.9 ppm as the isomerization product of **4** in addition to two signals at δ-44.0 and δ-44.4 ppm. Isomers were formed by the nucleophilic attack of the hydroxyl group of serine side chain. The isomerization could be reduced by controlling the reaction temperature. Moreover, it was supposed that in terms of the formation of pentacoordinated phosphorus intermediate, the formation rate constant of histidine was 2.53 times faster than that of serine because of the catalysis of imidazole ring [20]. Therefore, we hypothesized that the formation mechanism of His–His dipeptide was due to the competitive reaction of *O*,*O*-phenylenephosphoryl group between pentacoordinated phosphorus intermediates **4** and **7** as shown in Figure 5.3 [21].

In summary, there was a competition reaction between the formation of the pentacoordinted phosphorus compounds **4** and **7**. They could be attacked by *N*,*N*-bis (trimethylsilyl) amino acid trimethylsilyl ester. For the purpose of verifying the mechanism further, the seryl-phenylalanine dipeptide and the seryl-asparagine dipeptide have been synthesized by this method. Similarly, in the reaction mixture there were competition products such as phenylalanyl–phenylalanine dipeptide or aspartyl–asparagine dipeptide that were identified by the ESI-MS/MS, respectively. Therefore, it was proposed

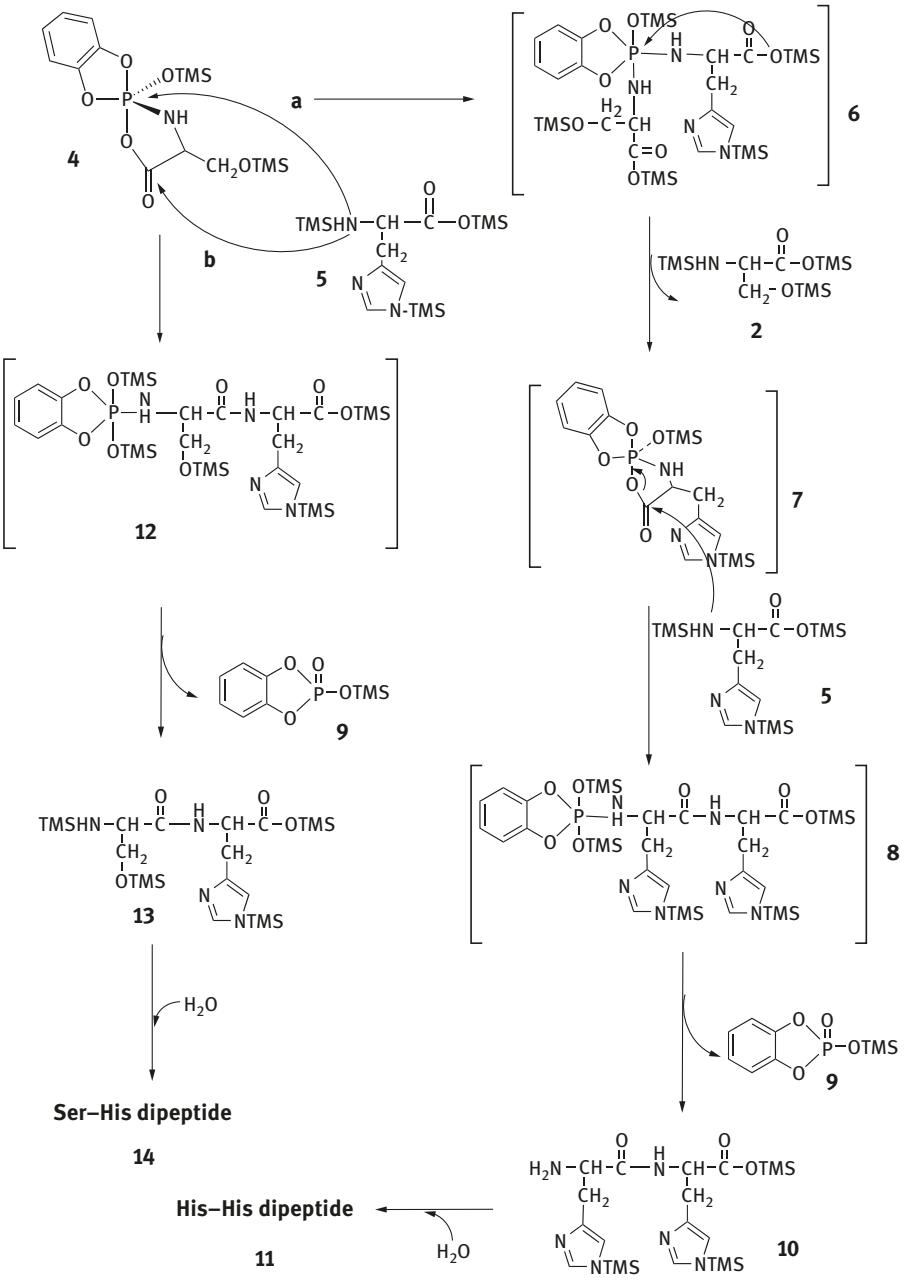


Figure 5.3: The proposed phosphoryl transfer mechanism for His–His formation.

that there was the ligand exchange of pentacoordinated phosphorus compound in the amino acid self-assembly reaction by the activation of phosphorous reagent.

5.3 The transfer reaction of *N*-(*O*,*O*-diisopropyl) phosphoryl between different amino acids

To investigate the intermolecular phosphoryl transfer between serine and histidine residues, the intermolecular phosphoryl transfer reaction of *O*-trimethylsilyl-*N*-(*O*,*O*-diisopropyl) phosphoryl serine trimethylsilyl ester (DIPP-Ser) **16** was studied. The reaction of DIPP-Ser and *N*,*N*-bis(trimethylsilyl) histidine trimethylsilyl ester **5** was performed. Equimolar amounts of **16** and **5** were mixed and reacted at room temperature. After three days, the reaction products were acidified with 5 M HCl. The normal product was the DIPP–Ser–His dipeptide, and there should be DIPP–His–Ser or DIPP–His–His if the phosphoryl group transferred from serine to histidine residue. The crude products in water phase were detected by HPLC-ESI-MS/MS. It was found that His–His dipeptide was produced (Figure 5.4).

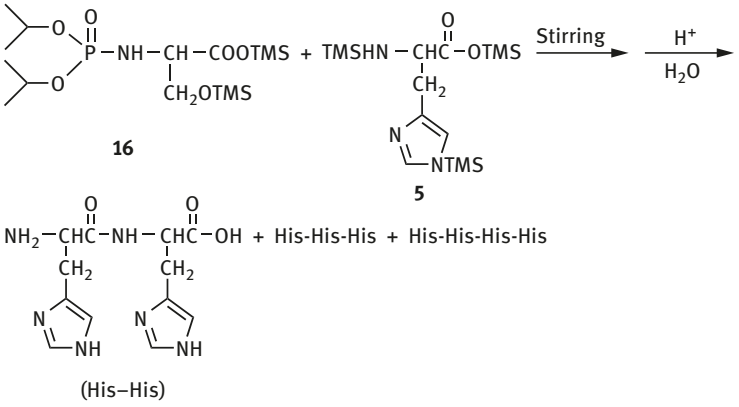


Figure 5.4: The phosphoryl transfer reaction between DIPP–Ser and histidine.

The HPLC-ESI-MS tracking experiment showed that the di-, tri-, tetrapeptides self-assembled from DIPP–His were produced, except for hydrolysis product, histidine (*m*/*z* 156). And no peptide from DIPP–Ser was produced, for example, Ser–Ser and Ser–his. The above results suggest that the intermolecular phosphoryl transfer takes place from *N*-(*O*,*O*-diisopropyl) phosphorylserine to *N*-histidine trimethylsilyl ester. The intermolecular phosphoryl transfer mechanism was proposed as in Figure 5.5 [22].

As shown in Figure 5.5, the basicity of histidine is stronger than that of serine at the start of the reaction. That means the nucleophilicity of nitrogen terminal

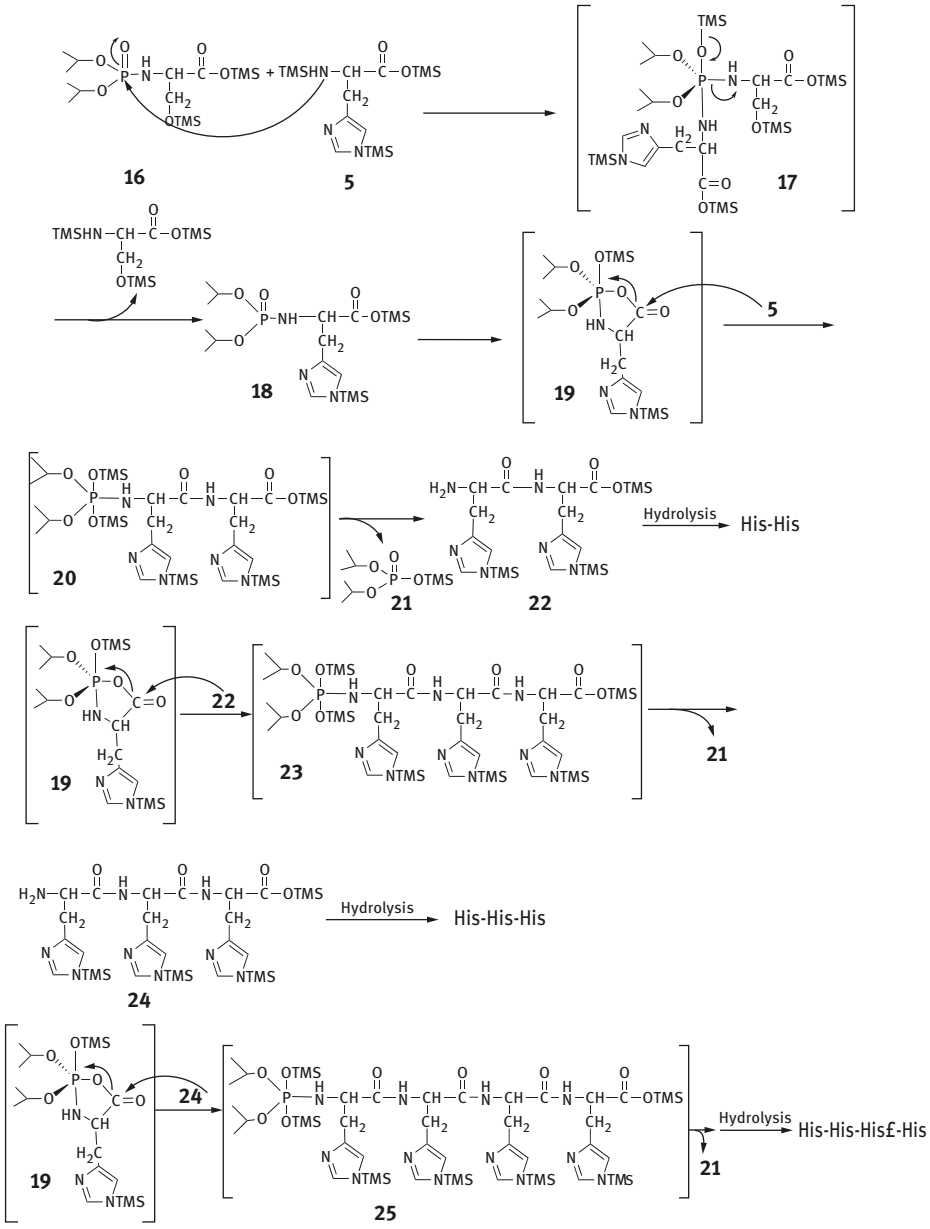


Figure 5.5: The proposed mechanism of intermolecular diisopropylphosphoryl transfer reaction.

of histidine is larger than that of serine. Therefore, the amino group of *N*,*N*-bis (trimethylsilyl) histidine trimethylsilyl ester **5** attacked the phosphoryl group of *O*-trimethylsilyl-*N*-(*O*,*O*-diisopropyl) phosphoryl serine trimethylsilyl ester **16**, and then by intermolecular phosphoryl transfer, *N*-trimethylsilyl-*N*-(*O*,*O*-diisopropyl)

phosphoryl histidine trimethylsilyl ester **18** was obtained. It was proposed that the transfer reaction went through pentacoordinated phosphorus intermediate **17**. Then histidine residue of **18** was activated through the formation of carboxylic-phosphoric mixed anhydride **19**, which could be considered as the model of amino acyl-tRNA. And it could be coupled with **5** to produce the His–His, His–His–His, oligopeptides sequentially.

In order to verify the formation of the pentacoordinated phosphorus intermediate, we used DIPP–His protected carboxyl group by methyl ester to react with DIPP–Ser. Because the carboxyl of histidine had undergone esterification, it could prevent the formation of pentacoordinated phosphorus activation intermediate. Therefore, the generation of dipeptide His–His was also prevented. It was expected that the reaction would stay in the step of phosphoryl histidine methyl ester and would not continue to generate dipeptide. The reaction is shown in Figure 5.6. The HPLC-ESI-MS tracking experiment showed that, due to the formation of DIPP-His-OMe after the phosphoryl transfer reaction, DIPP-His-OMe would not generate pentacoordinated phosphorus intermediate. Consequently, no dipeptide was produced, and the reaction stayed in the step of DIPP-His-OMe stage. The result of this experiment indicated that the intermolecular phosphoryl transfer from N-terminal of serine to that of histidine occurred, and the reaction finally stayed in DIPP-His-OMe stage.

5.4 The studies on the phosphoryl transfer reaction between different types of amino acids

Based on the studies of the intermolecular phosphoryl transfer reaction between DIPP-Ser and His, the phosphoryl transfer reaction between different types of phosphoryl amino acids with another amino acid was subsequently investigated. Considering the classification of amino acids and the function of some amino acids (serine, threonine, histidine and aspartic acid) in the phosphorylation and dephosphorylation process in vivo protease catalytic protein, we investigated the phosphoryl transfer reaction between some representative phosphoryl amino acids with different types of amino acid in ammonium acetate buffer solution.

According to the acidity or basicity of 20 common amino acids, the phosphoryl amino acids and amino acids are divided into three categories: acidic, neutral and basic amino acids. We designed a series of crossover experiment in ammonium acetate buffer between basic, acidic, neutral phosphoryl amino acids and amino acids, respectively. The reaction results were tracked and detected by ESI-MS/MS analysis, and the reaction mixture was separated and identified by HPLC-ESI-MS/MS. The results showed that in the reaction of the phosphoryl amino acids (P-aa1) and other amino acids (aa2) reaction, only when the basicity of aa2 was stronger than that of aa1, namely N-terminal nucleophilicity of aa2 was greater than that of aa1,

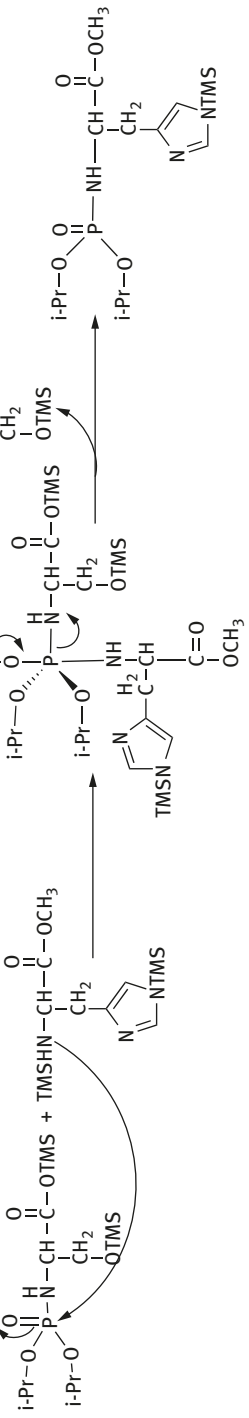


Figure 5.6: The phosphoryl transfer between DIPP-Ser and histidine methyl ester.

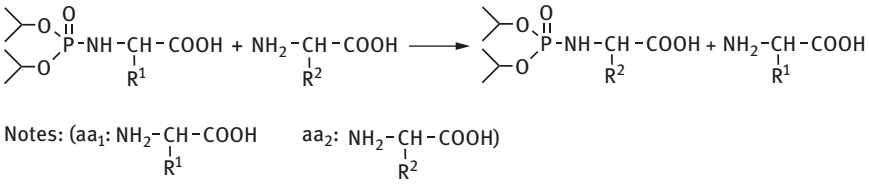


Figure 5.7: The phosphoryl transfer between DIPP-AA and another type of amino acid.

the phosphoryl group could transfer from aa1 to aa2. The reaction was shown in Figure 5.7. And it was found that the stronger the basicity of amino acids aa2 is, the more competitive of the phosphoryl group it has, and then the phosphoryl transfer reaction occurred more easily.

In order to further validate the above conclusion, the phosphoryl transfer reaction of the alkaline diamine compounds with stronger basicity (such as 1,6-hexamethylendiamine and 1,3-propane diamine) and phosphoryl amino acids was investigated. The results showed that the phosphoryl group was easily transferred to the diamine compounds and generated the phosphorylation diamine compounds. The N-terminal nucleophilicity of diamine compounds was greater than that of amino acids, so that the rate of phosphoryl transfer reaction was increased. The reaction was detected and identified by HPLC-ESI-MS/MS analysis. Moreover, the phosphoryl transfer reaction product was separated and identified by NMR [23].

The phosphoryl transfer reaction through pentacoordinated phosphoryl intermediate has been proposed by some direct or indirect experimental proof. The effect of the basicity of amino acid on the phosphoryl transfer was also explored. However, the phosphoryl transfer reaction mechanism and how to adjust or control the transfer reaction through pentacoordinated phosphoryl amino acids need to be further studied.

References

[1] Li YF, Wang Q, Zhao YF et al. The Chemical Properties of N-(*O,O*-diisopropyl) Phosphoryl-arginine[J]. *Phosphorus Sulfur Silicon*, 1995, 106: 131–136.

[2] Li YM, Zhang DQ, Zhang HW et al. β-Carboxyl Catalytic Effect of N-Phosphoryl Aspartic Acid[J]. *Bioorg. Chem.*, 1992, 20: 285–295.

[3] Li YC, Zhao YF, Synthesis and Properties of N-diisopropyloxyphosphoryl Cystein[J]. *Phosphorus Sulfur Silicon*, 1991, 60: 233–237.

[4] Ma XB, Zhao Y F, Phosphoryl Group Paticipation in the Reactions of N-Phosphoryldipeptide Acids[J]. *Phosphorus Sulfur Silicon* 1992, 66: 107–114.

[5] Li YM, Jin Y, Li YC, et al. Phosphorylation of Tyrosine and the Reactivity of Dialkyl Phosphites with Tyrosine[J]. *Phosphorus Sulfur Silicon*, 1995, 101: 141–147.

[6] Fu H, Li ZL, Zhao YF et al. Oligomerization of *N,O*-bis(trimethylsilyl) Amino Acids Mediated by O-phenylenephosphochloridate[J]. *J. Am. Chem. Soc.*, 1999, 121: 291–295.

[7] Li YM, Yin YW, Zhao YF, Phosphoryl Group Participation Leads to Peptide Formation from N-phosphoryl amino acids[J]. *Int. J. Peptide Protein Res.*, 1992, 39: 375–381.
[8] Zhao YF, Ju Y, Li YM, et al. Self-activation of N-phosphoamino acids and N-phosphodipeptides in oligopeptide formation[J]. *Int. J. Peptide Protein Res*., 1995, 45: 514–518.
[9] Zhou WH, Ju Y, Zhao YF, et al. Simultaneous Formation Peptides and Nucleotides from N-phosphothreonine[J]. *Origins of Life and Evolution of the Biosphere*, 1996, 26: 547–560.
[10] Xue CB, Yin YW, Zhao YF, Studies on Phosphoserine and Phosphothreonine Derivatives[J]. *Tetrahedron Lett.*, 1988, 29: 1145.
[11] Zhao YF, Cao PS, Phosphoryl Amino acids: Common Origin for Nucleic Acids and Protein[J]. *J. Biol. Phys.*, 1994, 20: 283–287.
[12] Zhao YF, Cao PS, Basic Models of Chemical Evolution of Life: The Minimum Evolving System. In: Chela-Flores J, Raulin F. eds. *Chemical Evolution: Physics of the Origin and Evolution of Life*[M]. Netherlands: Kluwer Academic Publishers, 1996, 279–285.
[13] Blatter W A, Knowles J R, Stereochemical Course of Glycerol Kinase, Pyruvate Kinese, and Hexokinase: Phosphoryl Transfer from Chiral [γ(s)-^{16}O, ^{17}O, ^{18}O]ATP[J]. *J. Am. Chem. Soc.*, 1979, 101: 510–511.
[14] Fersht A. (Du Jinzhu et al. Translate), Structure and Mechanism of Enzyme [M], 1991, Peking University Press, Beijing, China, 253.
[15] Lukat GS, Stock JB, et al. Roles of the Highly Conserved Aspartate and Lysine Residues in the Response Regulator of Bacterial Chemotaxis[J]. *J. Biol. Chem*., 1991, 266: 8348–8354.
[16] Qninn DM, Acetychollnesterase: Enzyme Structure, Reaction Dynamics, and Virtual Transition States[J]. *Chem. Rev.,* 1987, 87: 955–979.
[17] Wen JK, *Chemistry of Life*[M], 1993, 13: 15.
[18] Lu K, Tu GZ, Guo XF, et al. Structure and Isomerization of *O*,*O*-Phenylene Pentacoordinated Phosphoryl Serine[J]. *J. Mol. Struct.*, 2002, 610: 65–72.
[19] Lu K, Liu Y, Zhou N, et al. Synthesis of Seryl-histidine Dipeptide from Pentacoordinated Phosphoryl Serine[J]. *J. Tsinghua Univ. (Nat Sci Edn*), 2001, 41: 83–86.
[20] Wang Qian, Study on synthesis and reaction mechanism of amino acid pentacoordinated phosphorus compounds and polyene [D]. Doctoral dissertation, Tsinghua University, 1997.
[21] Cao SX, Niu MY, Su YQ et al. Ligand Exchange Between pentacoordinated Phosphoryl Serine and Histidine Compounds[J]. *Chinese J. Chem.*, 2003, 21: 1647–1651.
[22] Su YQ, Niu MY, Cao SX, et al. Intermolecular Phosphoryl Transfer Between Serine and Histidine Residues[J]. *Chinese Chem. Lett.*, 2004, 15: 1289–1291.
[23] Su Yuqian, Study on Intermolecular Phosphoryl Transfer Reaction[D]. Master dissertation, Zhengzhou University, 2005.

6 The research progress of chiral pentacoordinate spirophosphoranes with bis-α-amino acid bonds

On the basis of the long-range research of pentacoordinate phosphorus compounds in our laboratory, the stereochemistry of chiral bisamino acyl pentacoordinate spirophosphoranes and its progress with time across the world were studied. These compounds are arranged in a distorted TBP (triangular bipyramid/trigonal bipyramid) geometry. Two nitrogen atoms and a hydrogen atom form an equator plane and two oxygen atoms are in apical positions. Based on these data, we adopted the nomenclature system for the coordination compound $[MX(AB)_2]$ (AB = heterobidentate ligand) with a TBP geometry to determine the absolute configuration of the phosphorus center. And then the absolute configurations of all compounds were correlated with solid-state circular dichroism (CD) and ^{1}H NMR (nuclear magnetic resonance) spectroscopy. The ^{1}H–^{1}H COSY (correlated spectroscopy) of these compounds identified an interaction between the P–H proton and the α-hydrogen of amino acids, thus establishing unusual coupling through four bonds for these species with $^4J_{\text{H-C-N-P-H}} = 2.4$ Hz.

6.1 The importance of pentacoordinate phosphorus compounds in biological processes

Pentacoordinate phosphorus compounds, especially the phosphorus alkanes, are a totally new and hot area for nearly 30 years, and play a crucial role not only in organic phosphorus chemistry but also in biochemistry. In organic phosphorus reaction, transformation of many intermediates of tri- and tetracoordinated phosphorus compounds results in a pentacoordinated structure as in the biomimetic chemistry process of *N*-phosphoryl amino acids [1]. This compound could automatically produce peptides [2–5], process the esterification [6] and transesterification [7] and furthermore set up the transfer of phosphorus group from N atom to O atom via pentacoordinate phosphorus intermediates or transition states [8]. Many biological mechanisms are predominated by this species because of the importance of phosphorus element in organism. The regulation of the activity of some enzymes by phosphorylation or dephosphorylation in signal conduction of living cells, the energy transfer process of adenosine triphosphate and the autogenous cutting of RNA have been chemically conducted by phosphoryl transfer process, particularly via pentacoordinate spirophosphoranes with amino acid bonds. It therefore makes perfect sense to study the stereochemistry of these compounds.

Coordinated phosphorus compounds are a new type of organophosphorus compounds discovered in the 1960s, which explore a new research field in chemical

https://doi.org/10.1515/9783110562552-006

biological organic phosphorus by its abundant connotation. And then this leads to high coordination, including six coordination of phosphorus and organic phosphorus chemical development. Frank H. Westheimer and Fausto Ramirez made a pioneering contribution in this area. In 1966, when Westheimer [9] studied the hydrolysis of the five-ring cyclophosphate, he proposed that the reaction mechanism was the five-coordinated transition state. This hypothesis has successfully explained experimental results of the hydrolysis mechanism of the five-ring phosphate ester inside or outside, and adopted by other scientists because it could also explain the mechanism of cyclic or noncyclic phosphate ester hydrolysis. The work of Westheimer laid down the status of phosphorane in the field of organic phosphorus chemistry. Ramirez [10] has established a set of rules for the formation of cyclic phosphorus alkanes, and summarized the structure characteristics of pentacoordinated phosphorus compounds and the synthesis method of cyclic pentacoordinated phosphorus compounds. Two principles such as apicophilicity and pseudorotation decided the place of their coordination groups in "a" or "e" direction in the TBP configuration of phosphorus alkanes according to the structural analysis of a large number of stable pentacoordinated phosphorus compounds [11]. These two basic principles and also some other forces (e.g. hydrogen bond, van der Waals force and other noncovalent bond forces) codetermine the space arrangement of each of the coordinating groups in the phosphorus alkane.

To investigate the stereochemistry of chiral bisamino acyl pentacoordinate spirophosphoranes systematically, 16 isomers were synthesized and their absolute configurations were characterized by NMR, single-crystal X-ray diffraction, CD and the theoretical calculation of quantum chemistry.

6.2 The synthesis method of chiral pentacoordinate spirophosphoranes

Some literatures report the synthesis and property of pentacoordinated phosphorus compounds containing the fragment of amino acid (see Figure 6.1).

Compound **1** was synthesized by Gololobov [12] from *O*,*O*′- phenylenephenyldiphosphite and diazo carboxylic acid. Zaloom [13] synthesized compound **2** using the similar method. One novel pentacoordinate spirophosphorane **3** from amino acids was synthesized by Add-Ellah [14]. A lot of research is also done on this field by French scientists A. Munozhe and B. Garrigues [15, 16]. Their synthesis method and reaction behavior was optimized, as well as their stereostructure was determined. After synthesizing compounds **4**, **5** and **6**, the stereochemistry properties of compound **7** and dynamic transformation between it and its corresponding tetracoordinated phosphorus compounds were systematically studied and its barrier energy was determined. How the compound **7** could react with

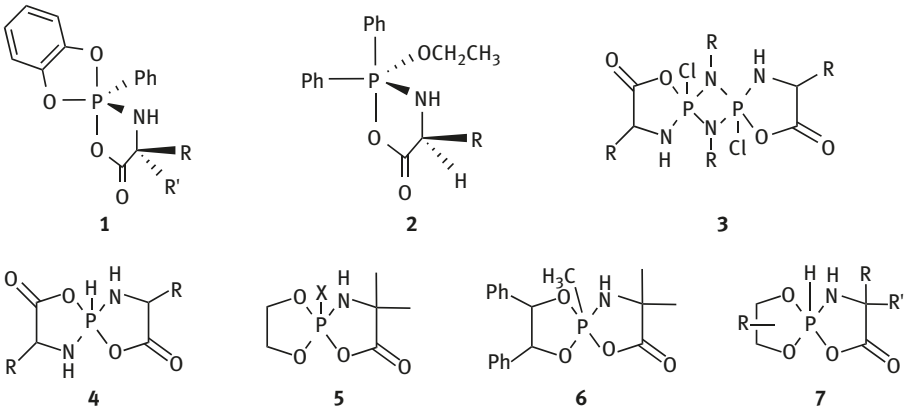


Figure 6.1: Pentacoordinate phosphorus compounds.

water, alcohol and base was followed up and the consideration of hexacoordinated phosphorus compound from compound **7** was also made by these two scientists [17].

The pentacoordinate spirophosphoranes described earlier consist of only one amino acid. Indeed, it is the rarely reported synthesis of pentacoordinate spirophosphoranes with bis-α-amino acid bonds (only glycine and alanine). Pentacoordinate spirophosphoranes with bis-α-amino acid bonds were first synthesized by Garrigue [15] in 1970 (see Figure 6.2).

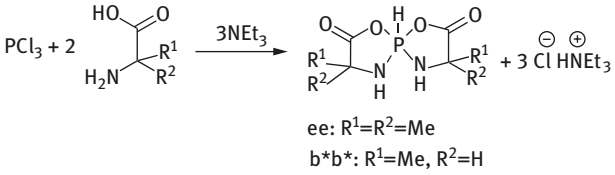
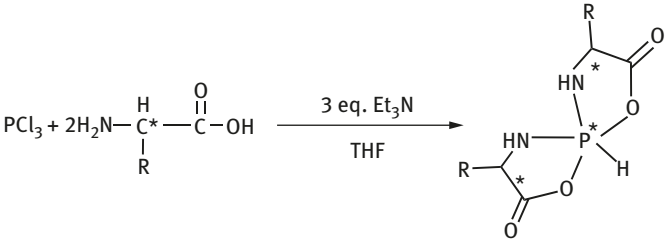


Figure 6.2: The synthesis method of Garrigue.

In our laboratory, special α-amino acid with no active functional group on the side chain and also its N and O atoms protected by Trimethylsilyl chloride (TMSCl) could form monoamino acid-chelated pentacoordinate phosphorus intermediate, which would produce 2–*n* oligomeric peptide via self-assembly [4].

In order to further determine and study the absolute configuration of bisamino acyl pentacoordinate spirophosphoranes, 16 compounds of this kind were synthesized using Garrigue's method from the start materials – amino acids (see Figure 6.3) and purified by fast flash column. The ^{31}P NMR chemical shift of all of the compounds is listed in Table 6.1 and the integral area of corresponding spectral peak was also calculated which was in direct proportion to the ratio of the products.



L-Valine	3a, 3b
D-Valine	4a, 4b
L-Leucine	5a, 5b
D-Leucine	6a, 6b
L-Phenylglycine	7a, 7b
D-Phenylglycine	8a, 8b
L-Phenylalanine	9a, 9b
D-Phenylalanine	10a, 10b

Figure 6.3: The synthesis of chiral bisamino acyl pentacoordinate spirophosphoranes.

Table 6.1: The ^{31}P NMR chemical shift and ratio of chiral bisamino acyl pentacoordinate spirophosphoranes.

Amino acid	Products	Chemical shift[a]	Ratio (%)[b]
L-Val	**3a**	−64.80	38
	3b	−61.68	62
D-Val	**4a**	−64.79	35
	4b	−61.70	65
L-Leu	**5a**	−64.50	46
	5b	−63.83	54
D-Leu	**6a**	−64.54	49
	6b	−63.83	51
L-PhGly	**7a**	−63.73	47
	7b	−61.57	53
D-PhGly	**8a**	−63.50	42
	8b	−61.34	58
L-Phe	**9a**	−63.03	55
	9b	−60.03	45
D-Phe	**10a**	−63.15	44
	10b	−60.09	56

[a]^{31}P NMR data were obtained in the dimethylsulfoxide (DMSO) solution. Once the reaction is completed, the tetrahydrofuran solvent was removed by rotary evaporation and the residue (diastereoisomers a and b) was completely dissolved in DMSO.
[b]The ratio of products was calculated from the integral area of ^{31}P NMR signals

6.3 Spectroscopic characterization of bisamino acyl pentacoordinate spirophosphoranes

In order to determine the absolute configuration of phosphorus center, 16 diastereoisomers of pentacoordinate spirophosphoranes were synthesized and isolated. Their spectroscopic property was studied systematically according to the compounds described below [18].

6.3.1 $^4J_{H\text{-}C\text{-}N\text{-}P\text{-}H}$

Herein, one set of diastereomers of chiral pentacoordinate spirophosphoranes derived from L-valine (or D-valine) was synthesized and separated, respectively (**3a, 3b, 4a, 4b**). The above compounds and four isomers were also characterized by ^{1}H NMR solution spectroscopy. With respect to the signals arising from the P–H function of these pentacoordinate spirophosphoranes, the epimeric **a/b** pairs of the enantiomers **3a/4a** and **3b/4b** show significantly different spectra (Figure 6.4) while the spectra of enantiomers **3a/4a**, likewise of **3b/4b**, are identical.

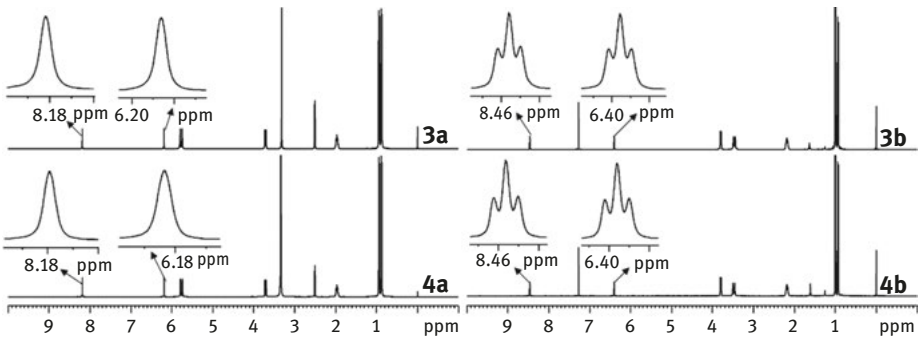


Figure 6.4: ^{1}H NMR spectra of compounds 3a–4b. (3a/4a: DMSO-d_6, δ_{PH} = 7.18, d, $^1J_{PH}$ = 798.5 Hz; 3b/4b: $CDCl_3$, δ_{PH} = 7.43, dt, $^1J_{PH}$ = 824.4 Hz, $^4J_{HH}$ = 2.4 Hz).

Whether the proton bound to nitrogen is deuterated or not, the proton bound to phosphorus in **3b** and **4b** gives a double triplet signal from coupling to phosphorus and to two magnetically equivalent protons (d = 7.43 ppm, dt, $^1J_{(H,P)}$ = 824.4 Hz, $^4J_{(H,H)}$ = 2.4 Hz). However, the P–H proton in **3a** and **4a** shows a doublet signal split only by phosphorus (d = 7.18 ppm, d, $^1J_{(H,P)}$ = 798.5 Hz). This phenomenon was further confirmed by the ^{1}H–^{1}H COSY of **3b** and **4b** identifying an interaction between the P–H proton and the α-hydrogen of valine, thus establishing unusual coupling through four bonds for these species. There is no such ^{1}H–^{1}H COSY effect for the epimeric **3a/4a** isomers. The ^{1}H NMR

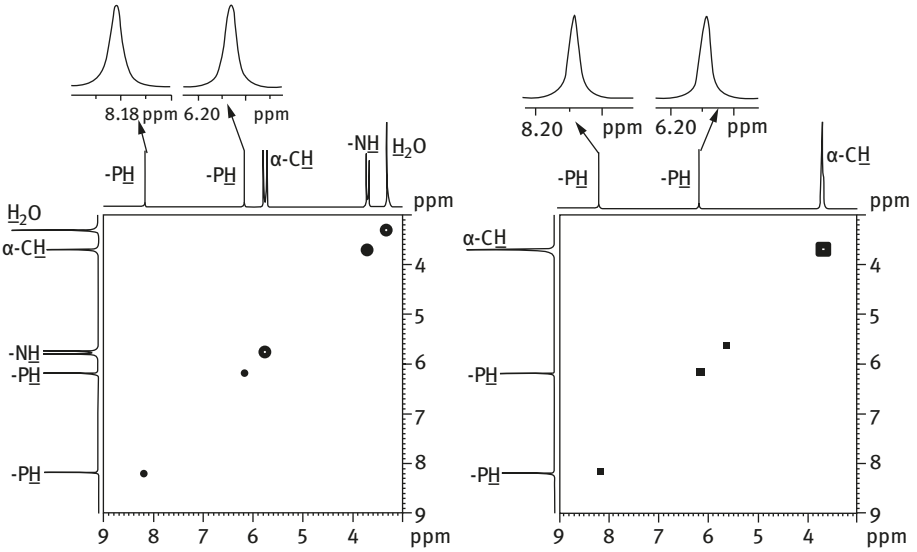


Figure 6.5: $^{1}H-^{1}H$ COSY of compound 3a ($\mathcal{A}_P$,S_C,S_C) [Solvent: DMSO-d_6 (left), DMSO-d_6 + D_2O (right)].

spectra of **3a–4b** are shown in Figure 6.4, and could support the conclusions from the CD spectra that each pair of **3a/4a** and **3b/4b** is enantiomeric, respectively (Figure 6.5).

This special phenomenon of coupling was further confirmed by the $^{1}H-^{1}H$ COSY of **3b/4b**. Whether the N–H proton was deuterated or not, the $^{1}H-^{1}H$ COSY identified an interaction between the P–H proton and the α-hydrogen of the amino acid establishing an unusual four-bond distance coupling for these cases (the $^{1}H-^{1}H$ COSY of **3b** in $CDCl_3$ and $CDCl_3$ + D_2O are shown in Figure 6.6). However, no such $^{1}H-^{1}H$ COSY effect was detected in the epimers (Figure 6.5).

Different chiral pentacoordinate spirophosphoranes derived from different amino acids have the same principle of NMR coupling (see Table 6.2).

It is obvious that the comparison of results showed certain regularity and all of the compounds could be divided as two parts: all the compounds in **"a"** configuration show a doublet signal split only by phosphorus and there is no such $^{1}H-^{1}H$ COSY effect for the epimeric **"a"** isomers. On the other hand, all the compounds in **"b"** configuration give a double triplet signal from coupling to phosphorus and to two magnetically equivalent protons and the $^{1}H-^{1}H$ COSY identified an interaction between the P–H proton and the α-hydrogen of the amino acid establishing an unusual four-bond distance coupling for these cases.

According to the X-ray structural analysis of these chiral pentacoordinate spirophosphoranes, the P–H proton and the α-hydrogen of the amino acid form a pentacyclic ring. Therefore, this interaction could be conducted by the sum of the coupling constants $^{4}J_{H\text{-}C\text{-}N\text{-}P\text{-}H}$ and $^{5}J_{H\text{-}C\text{-}C\text{-}O\text{-}P\text{-}H}$. Moreover, $^{5}J_{H\text{-}C\text{-}C\text{-}O\text{-}P\text{-}H}$ is separated by five single chemical bonds, which do not show any conjugation property or "W" shape orientation. So the coupling action of $^{5}J_{H\text{-}C\text{-}C\text{-}O\text{-}P\text{-}H}$ is quite weak. On the contrary, $^{4}J_{H\text{-}C\text{-}N\text{-}P\text{-}H}$ dominates this cou-

Table 6.2: The coupling constant of $^{4}J_{\text{H-C-N-P-H}}$ for all pentacoordinate spirophosphoranes.

Compounds	Solvent[a]	$^{4}J_{\text{H-C-N-P-H}}$ (Hz)	^{1}H–^{1}H COSY[b]
3a, 4a	DMSO-d_6 or DMSO-d_6 + D_2O	Not apparent split	×
3b, 4b	$CDCl_3$ or $CDCl_3$ + D_2O	2.4	√
5a, 6a	$CDCl_3$ or $CDCl_3$ + D_2O	Not apparent split	×
5b, 6b	$CDCl_3$or $CDCl_3$ + D_2O	2.4	√
7a, 8a	DMSO-d_6 or DMSO-d_6 +D_2O	Not apparent split	×
7b, 8b	DMSO-d_6 or DMSO-d_6 + D_2O	2.5	√
9a, 10a	DMSO-d_6 or DMSO-d_6 + D_2O	Not apparent split	×
9b, 10b	DMSO-d_6 or DMSO-d_6 + D_2O	2.4	√

[a]D_2O as the solvent to substitute the active hydrogen of NH.
[b]"×"means no ^{1}H–^{1}H COSY effect;"√"means ^{1}H–^{1}H COSY effect could be detected.

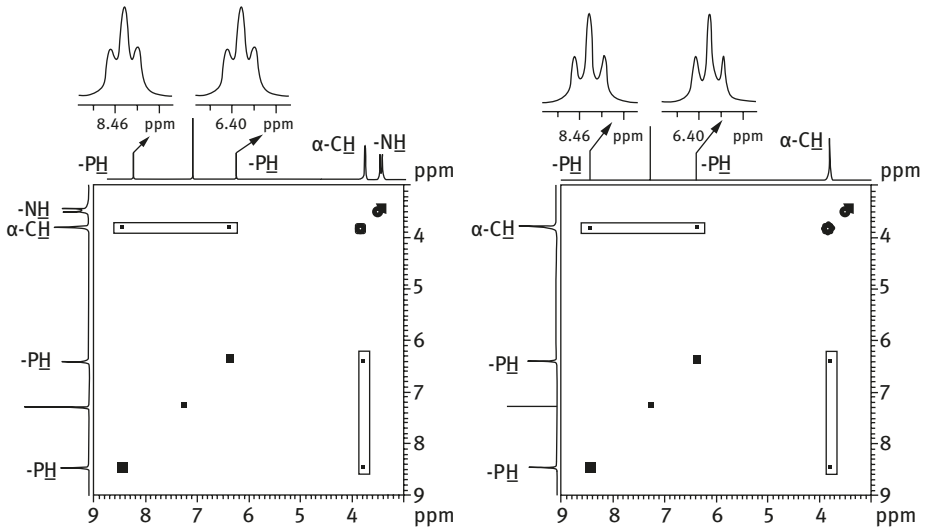


Figure 6.6: ^{1}H–^{1}H COSY of compound 3b (Δ_P,S_C,S_C). [Solvent: $CDCl_3$ (left), $CDCl_3$ + D_2O (right)].

pling effect. The configuration of phosphorus of pentacoordinate spirophosphoranes is a distorted TBP geometry. Two nitrogen atoms and a hydrogen atom form an equator plane and the angle of P–N–C is close to the ideal angle of 120°, which indicates that the chemical bond of nitrogen atom constructs a structure that is close to the plane, but not the TBP geometry. And the transfer of electrons from nitrogen atom to phosphorus atom results in the P–N bond to have the property of a double bond. Hence, H-C-N=P-H is similar to the *ene*-type structure. That is the reason why $^{4}J_{\text{H-C-N-P-H}}$ is more rational and important than $^{5}J_{\text{H-C-C-O-P-H}}$ in this structure. This phenomenon was confirmed by

theoretical calculation in quantum chemistry. For diastereoisomers, the difference between the extensional orientation of P–H and C–H bonds results in the difference in the ^{1}H–^{1}H COSY, and the coupling constant of $^{4}J_{H\text{-}C\text{-}N\text{-}P\text{-}H}$ was 2.5 or 0.5 Hz.

6.3.2 $^{1}J_{P\text{-}X}$

The P–N bond in the X-ray structure of these chiral pentacoordinate spirophosphoranes has the property of a double bond. However, the configuration of phosphorus of pentacoordinate spirophosphoranes is a distorted TBP geometry, and two nitrogen atoms and a hydrogen atom form an equator plane. Our experimental data show that the coupling constant of P–H and P–N bonds is quite different between tetracoordinate phosphorus compounds (see Figure 6.7) and pentacoordinate spiro-

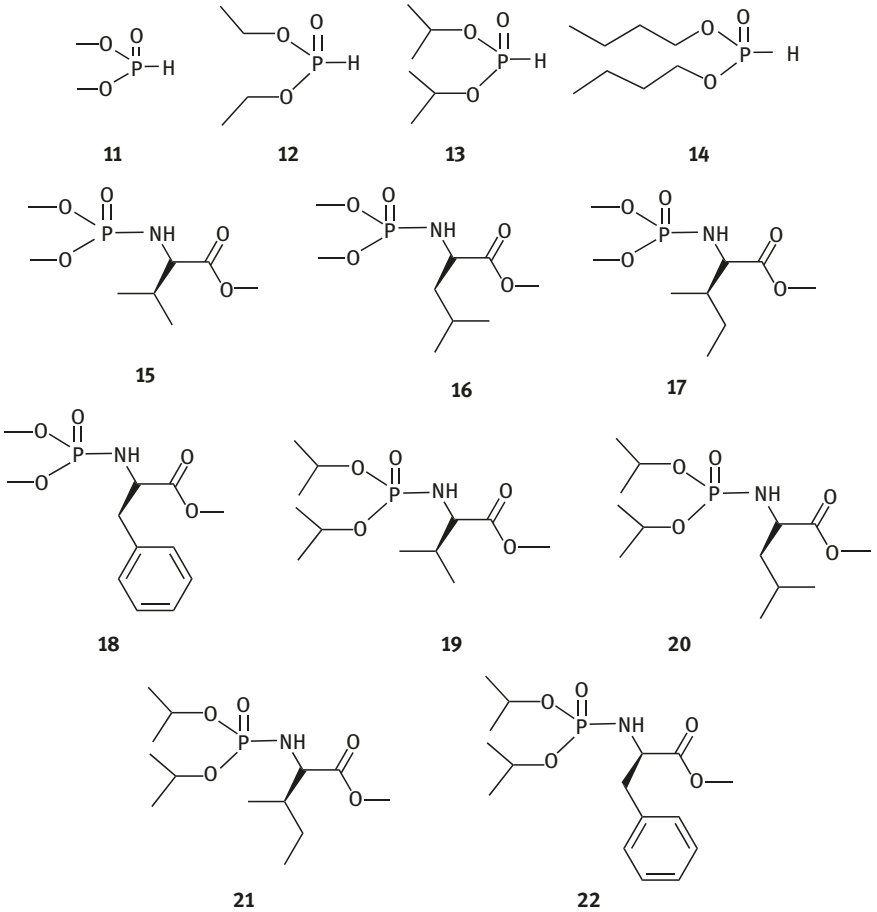


Figure 6.7: The tetracoordinate phosphorus compounds 11–22.

Table 6.3: The coupling constant of $^{1}J_{P\text{-}H}$ for compounds (3a–10b) and (11–14).

Compounds	$^{1}J_{P\text{-}H}$ (Hz)	Compounds	$^{1}J_{P\text{-}H}$ (Hz)	Compounds	$^{1}J_{P\text{-}H}$ (Hz)
3a[(a)]	798.5	**7a**[a]	821.0	**11**[(b)]	698.6
3b[(b)]	824.4	**7b**[a]	819.3	**12**[(b)]	692.6
4a[(a)]	798.6	**8a**[a]	821.0	**13**[(b)]	687.8
4b[(b)]	824.4	**8b**[a]	819.4	**14**[(b)]	692.0
5a[(a)]	810.9	**9a**[a]	804.9		
5b[(a)]	816.0	**9b**[a]	810.2		
6a[(a)]	810.9	**10a**[a]	805.0		
6b[(a)]	816.0	**10b**[a]	810.2		

[a]Solvent: DMSO-d_6
[b] solvent: $CDCl_3$

Table 6.4: The coupling constant of $^{1}J_{P\text{-}N}$ for compounds (3a–10b) and (11–14).

Compounds	$^{1}J_{P\text{-}N}$ (Hz)	Compounds	$^{1}J_{P\text{-}N}$ (Hz)	Compounds	$^{1}J_{P\text{-}N}$ (Hz)
3a[a]	35.0	**7a**[a]	32.9	**15**[b]	42.4
3b[b]	32.7	**7b**[a]	31.0	**16**[b]	42.3
4a[a]	34.9	**8a**[a]	32.9	**17**[b]	42.3
4b[b]	32.7	**8b**[a]	30.9	**18**[b]	42.0
5a[a]	32.4	**9a**[a]	34.5	**19**[b]	40.1
5b[a]	28.4	**9b**[a]	33.7	**20**[b]	40.8
6a[a]	32.3	**10a**[a]	34.6	**21**[b]	40.7
6b[a]	28.6	**10b**[a]	33.6	**22**[b]	41.0

[a]Solvent: DMSO-d_6
[b]solvent: $CDCl_3$.

phosphoranes (see Tables 6.3 and 6.4). The experimental results showed that $^{1}J_{P(IV)\text{-}H}$ (~700 Hz) < $^{1}J_{P(V)\text{-}H}$ (~800 Hz); but $^{1}J_{P(IV)\text{-}N}$ (~41 Hz) > $^{1}J_{P(V)\text{-}N}$ (~32 Hz).

The factors influencing the coupling constant ^{1}J are quite simpler than ^{3}J and other long-range coupling. It does not involve the angle of space, but is only relevant to the chemical bond property of two atoms producing spin–spin coupling. In order to elucidate our experimental results, quantum chemistry calculation of all these compounds was done. The geometry was optimized by DFT/B3LYP/6-31+G(d,p), and DFT/B3LYP/IGLO III was used to calculate the $^{1}J_{P\text{-}X}$.

$^{1}J_{P\text{-}X}$ project was summarized by Gorenstien [19] and the result shows that the Fermi contact spin–spin coupling dominates this coupling effect. Hence, the value of $^{1}J_{P\text{-}X}$ is affected by the electron density of *s* orbital in both P and N atoms between this bond. The larger the percentage of *s* orbital, the larger the value of coupling constant of $^{1}J_{P\text{-}X}$. Its empirical formula is

$$^{1}J_{P-X} = \frac{A\alpha_{p}^{2}\alpha_{x}^{2}}{1 + S_{p-x}^{2}} + B$$

where parameters A and B are constants; α_{p}^{2} and α_{x}^{2} represent the ratio of s orbital in P and X atoms. S_{p-x} is the overlapping factor of P–X bond.

With this equation in hand, we found that the phosphorus atom of tetracoordinate phosphorus compounds (see Figure 6.7, **11–22**) makes bonding with X atom by sp^3 hybridized orbitals. The structure of pentacoordinate spirophosphoranes is a distorted TBP geometry, and two nitrogen atoms and a hydrogen atom form an equator plane. And the hybrid orbital of central phosphorus is $sp^3d_{(z)}^2$ in this distorted TBP geometry when bonding with other atoms but is still sp^2 in the direction of equator plane.

In the P–H bond, the proton atom has only one electron on the s orbital, so the difference of $^{1}J_{P\text{-}H}$ between tetracoordinate phosphorus compounds and pentacoordinate spirophosphoranes is from hybrid form of phosphorus atom. In the tetracoordinate structure, phosphorus atom is bonding as sp^3 orbital, so s% is 25%. Whereas in the TBP geometry of pentacoordinate spirophosphoranes, phosphorus atom is bonding as sp^2 orbital in the direction of equator plane, so s% is 33.3% which is bigger than that of tetracoordinate phosphorus compounds. That is the reason why $^{1}J_{P(IV)\text{-}H}$ (~700 Hz) < $^{1}J_{P(V)\text{-}H}$ (~800 Hz).

In the P–N bond, unlike simple proton atom, the nitrogen atom has electron not only on the s orbital, but also on the p orbital. So the difference of $^{1}J_{P\text{-}H}$ between tetracoordinate phosphorus compounds and pentacoordinate spirophosphoranes is from both the hybrid forms of phosphorus and nitrogen atoms. In the tetracoordinate structure, phosphorus and nitrogen atoms are bonding as sp^3 orbital, so both of s% is 25%. Whereas in the TBP geometry of pentacoordinate spirophosphoranes, phosphorus and nitrogen atoms are bonding as sp^2 orbital and both of the s% is 33.3% which is bigger than that of tetracoordinate phosphorus compounds. According to the theory of Gorenstien, the coupling constant of $^{1}J_{P\text{-}N}$ of pentacoordinate spirophosphoranes is bigger than that of tetracoordinate phosphorus compounds. But the experimental data are the opposite. The quantum chemistry calculation result elucidates that the P–N bond in pentacoordinate spirophosphoranes displayed a nature of double bond decreasing the percent of s orbital to about 25%, which resulted in the decrease of the $^{1}J_{P\text{-}N}$ value.

6.3.3 X-ray

The absolute configurations of central phosphorus atom (Δ and Λ) of a pair of enantiomers **3a** and **4a** were determined by X-ray diffraction analysis in our laboratory, and the other pairs of enantiomers could be correlated with solid-state CD and ^{1}H NMR spectra and named by the nomenclature rule in coordination stereochemistry (see Figure 6.8).

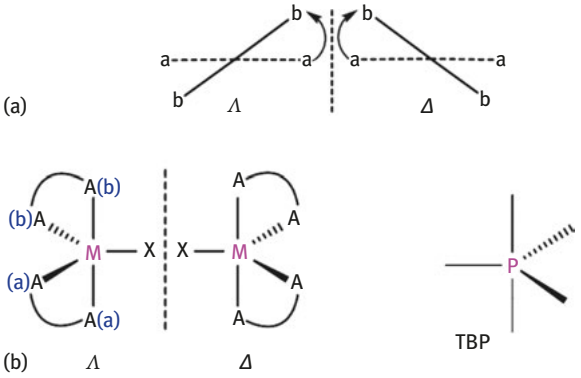


Figure 6.8: The nomenclature (*Λ* and *Δ*) of possible chiral configurations of pentacoordinated complexes $[MX(AB)_2]$ in TBP geometry.

Apparently, the absolute configurations of **3a–6b** could also be correlated with the ^{1}H NMR signal. For the pentacoordinate spirophosphoranes derived from L-amino acids, when the proton bound to phosphorus gives a doublet signal, the absolute configuration of the phosphorus atom could be assigned as $\mathbf{\Lambda}_P$, and when it shows a double triplet signal, the absolute configuration of phosphorus could be assigned as $\mathbf{\Delta}_P$. Likewise, for the pentacoordinate spirophosphoranes derived from D-amino acids, when the proton bound to phosphorus gives a doublet, the absolute configuration of phosphorus could be assigned as $\mathbf{\Delta}_P$, and when the proton bound to phosphorus shows a double triplet, the absolute configuration of phosphorus could be assigned as $\mathbf{\Lambda}_P$. The absolute configurations of **3a–6b** were listed in Table 7.5.

The absolute configurations of these pentacoordinate spirophosphoranes were proved when the single-crystal structures of a pair of enantiomers **3a/4a** were characterized by X-ray diffraction analysis. Crystals of **3a** and **4a** suitable for X-ray analysis were obtained from acetone solutions. Oak Ridge Thermal-Ellipsoid Plot (ORTEP) structures of **3a/4a** are shown in Figure 6.9.

In order to identify the absolute configuration of the phosphorus center, we adopted the nomenclature system for a coordination compound $[MX(AB)_2]$ (AB = heterobidentate ligand), which can be applied to a TBP or square-pyramidal (S_P) geometry. In the TBP geometry, when a monodentate ligand X occupies one equatorial position, the chiral-at-metal configuration can be defined as $\mathbf{\Lambda}_P$ or $\mathbf{\Delta}_P$ (see Figure 6.8).

According to the above nomenclature, the X-ray diffraction analysis shows that the phosphorus center in **3a** should be assigned as $\mathbf{\Lambda}_P$, and both α-carbons of the amino acids are in an (*S*)-configuration. Thus, the absolute configuration of compound **3a** is ($\mathbf{\Lambda}_P$, S_C, S_C). Likewise, **4a** is ($\mathbf{\Delta}_P$, R_C, R_C). Since **3b/4b**, **5a/6a** and

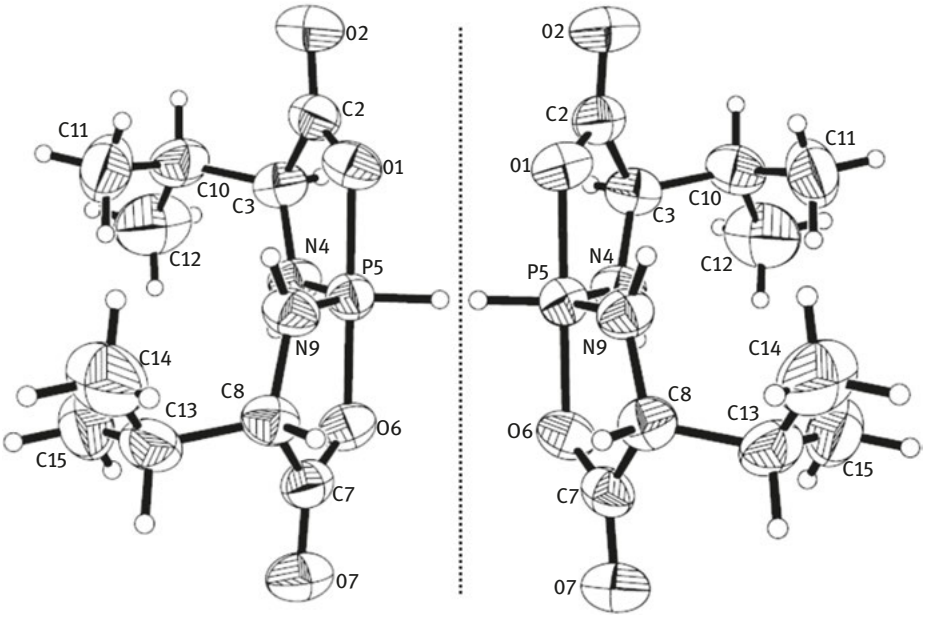


Figure 6.9: ORTEP drawing of compounds 3a (left) and 4a (right) with atom numbering scheme.

5b/6b could not be obtained as suitable single crystals for structure determination, their absolute configurations cannot be assigned directly.

The two compounds are arranged in a distorted TBP geometry. Two nitrogen atoms and a hydrogen atom form an equator plane and two oxygen atoms are in apical positions. The angles N(4)–P(5)–O(1), N(9)–P(5)–O(1), N(4)–P(5)–O(6) and N(9)–P(5)–O(6) around the phosphorus atom are close to 90°. However, the angles N(4)–P(5)–N(9) in 3a and 4a are not perfect with a deviation of 4.57° or 5.15° from the ideal angle of 120°, respectively. In the crystal lattice structures, the occurrence of N–H–O intermolecular hydrogen-bonding interactions led to the formation of a chain parallel to the *b*-axis and then the van der Waals interactions provide stability for the crystal structures. The veracity of absolute stereochemistry of **3a** and **4a** can be evaluated by the Flack parameters, which are 0.06(11) and 0.06(19), respectively. The crystal structure of **3a** was also obtained by Shu-Xia Cao. The crystal structure of **3a** or **4a** exhibits an endoconfiguration which looks like a "resting butterfly" with "wings" of two isopropyl groups of valine. From the aforementioned results, we can deduce that the noncrystalline enantiomers **3b** and **4b** may have an exoconfiguration which looks like a "resting moth." A similar crystal structure for the corresponding bis-alanine spirophosphorane has been obtained in our group.

Taken together, the absolute configuration of compounds (**3a, 4a, 7a, 8a, 7b, 8b, 9a, 10a**) was determined by X-ray diffraction analysis (see Table 6.5).

Table 6.5: The determination of absolute configuration of pentacoordinate spirophosphoranes.

Compounds	Configuration	Compounds	Configuration
3a	(Λ_P,S_C,S_C)	**4a**	(Δ_P,R_C,R_C)
7a	(Λ_P,S_C,S_C)	**8a**	(Δ_P,R_C,R_C)
7b	(Δ_P,S_C,S_C)	**8b**	(Λ_P,R_C,R_C)
9a	(Λ_P,S_C,S_C)	**10a**	(Δ_P,R_C,R_C)

6.3.4 Solid-state CD spectra

X-ray single crystal diffraction method is the most effective way to determine the absolute configuration of chiral molecule on the premise of obtaining a high-quality crystal of target molecule.

Since **3b/4b**, **5a/6a** and **5b/6b** could not be obtained as suitable single crystals for structure determination, their absolute configurations cannot be assigned directly.

In order to correlate the absolute configurations of each pair of isomers of **3b/4b**, **5a/6a** and **5b/6b**, their solid-state CD spectra were measured and are shown (Figures 6.10 and 6.11). The solid-state CD spectra show that **3a/4a**, **5a/6a**, **3b/4b** and **5b/6b** are indeed pairs of enantiomers. It is interesting to note that although **3a/3b** and **5a/5b** are synthesized from L-valine and L-leucine, respectively, they showed opposite Cotton effects. The same phenomenon was also found in **4a/4b** and **6a/6b**, which are derived from D-valine and D-leucine, respectively. It should be noted that the sign of the Cotton effects does not follow the chirality of the amino acid and the controlling factor for the asymmetry of these isomers **3a–6b** must be the chirality of the phosphorus center [20].

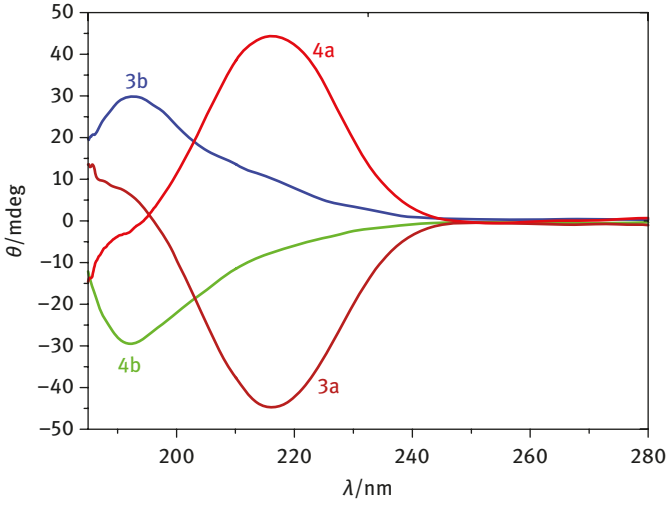


Figure 6.10: Solid-state CD spectra of compounds (3a–4b).

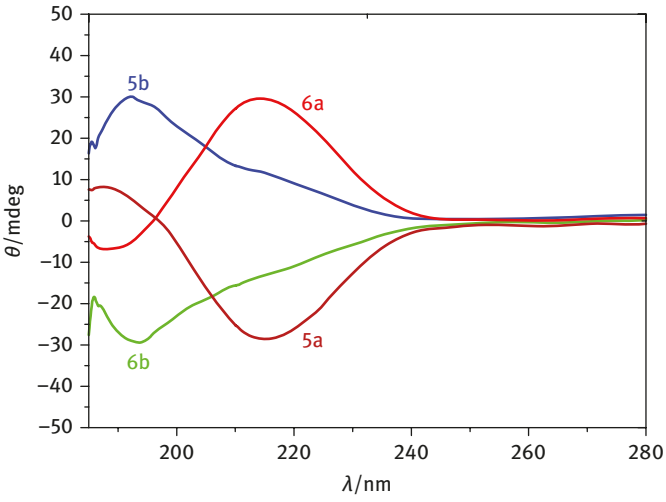


Figure 6.11: Solid-state CD spectra of compounds (5a–6b).

For example, **3a/3b** is derived from L-valine, and the absolute configurations of the two **α**-carbons are **S**, but they have opposite CD signs. Hence it is reasonable to conclude that they have opposite absolute configurations at the phosphorus atom. According to the absolute configuration of **3a** (**Λ_P, S_C, S_C**), we can thus assign the absolute configuration of **3b** as (**Δ_P, S_C, S_C**). Likewise, (**Λ_P, R_C, R_C**) is used to describe the absolute configuration of compound **4b**. The side chain of valine and leucine is not chromogenic and the difference of them is only $-CH_2$ unit, so the Cotton effects of **5a–6b** are similar to those of **3a–4b** and the absolute configurations of **5a–6b** can also be correlated and are listed in Table 6.6.

Table 6.6: Solid-state CD spectra of pentacoordinate spirophosphoranes and their correlation with absolute configuration.

Com pounds	CD signs	CONFIGURATION	Com pounds	CD signs	Configuration
3a*	–	(Λ_P,S_C,S_C)	**4a***	+	(Δ_P,R_C,R_C)
3b	+	(Δ_P,S_C,S_C)	**4b**	–	(Λ_P,R_C,R_C)
5a	–	(Λ_P,S_C,S_C)	**6a**	+	(Δ_P,R_C,R_C)
5b	+	(Δ_P,S_C,S_C)	**6b**	–	(Λ_P,R_C,R_C)
7a*	–	(Λ_P,S_C,S_C)	**8a***	+	(Δ_P,R_C,R_C)
7b*	–(205–225 nm) +(225–260 nm)	(Δ_P,S_C,S_C)	**8b***	+(205–225 nm) –(225–260 nm)	(Λ_P,R_C,R_C)
9a*	–	(Λ_P,S_C,S_C)	**10a***	+	(Δ_P,R_C,R_C)
9b	+	(Δ_P,S_C,S_C)	**10b**	–	(Λ_P,R_C,R_C)

The determination of absolute configuration of compounds **7a–8b**: The X-ray crystal structures of these four compounds were obtained and hence their absolute configuration could be determined directly. The solid CD spectra of 7**a**/**8a** and **7b**/**8b** are mirror-image symmetry, which demonstrate that they are indeed pairs of enantiomers (see Figure 6.12).

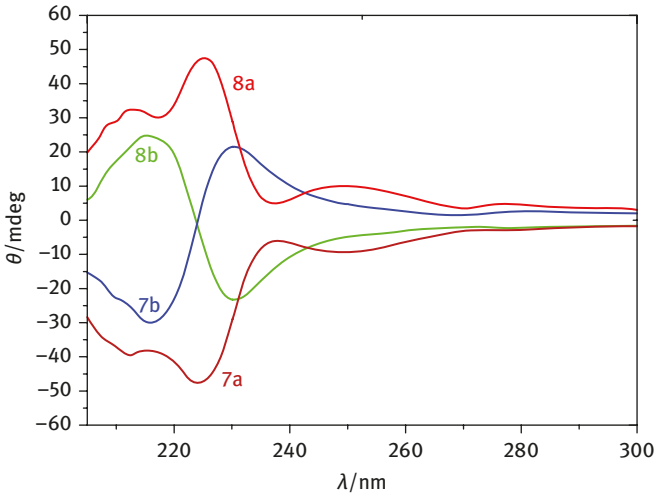


Figure 6.12: Solid-state CD spectra of compounds (7a–8b).

The CD spectra of **9a** and **10a** are virtually mirror images of each other (see Figure 6.13). Similarly, mirror-image symmetric CD spectra confirm the enantiomeric nature of **9b** and **10b**. Hence it is established that **9a/10a** and **9b/10b** are two pairs of enantiomers. Although both **9a** and **9b** are synthesized from L-phenylalanine, they have very different CD Cotton effects. Similarly, the same phenomenon is observed for **10a** and **10b**, both derived from D-phenylalanine. It is to be noted that the sign of the CD Cotton effect does not follow the chirality of the parent amino acid. The absolute configuration at the phosphorus center is the controlling factor for the optical asymmetry of the four isomers **9a–10b**. The absolute configuration of **9a** is ($\mathbf{\Lambda_P, S_C, S_C}$). Likewise, **10a** is (**ΔP, RC, RC**). Because **9b** and **10b** have not yet been crystallized, their absolute configurations cannot be assigned directly. However, since **9a** and **9b** are both derived from L-phenylalanine, the configurations of the two α-carbons are **S**. In addition, they have opposite CD signs, and therefore opposite configurations at phosphorus. We thus assign the configuration of **9b** as ($\mathbf{\Delta_P, S_C, S_C}$). Therefore, the configuration of **10b** is ($\mathbf{\Lambda_P, R_C, R_C}$).

As we know, we could determine the absolute configuration of some of our compounds by X-ray analysis. It is pity that we could not get the crystal structure for all of the compounds. Therefore, we have to derivate the absolute configuration of chiral atom of pentacoordinate spirophosphoranes by correlated solid CD spectra.

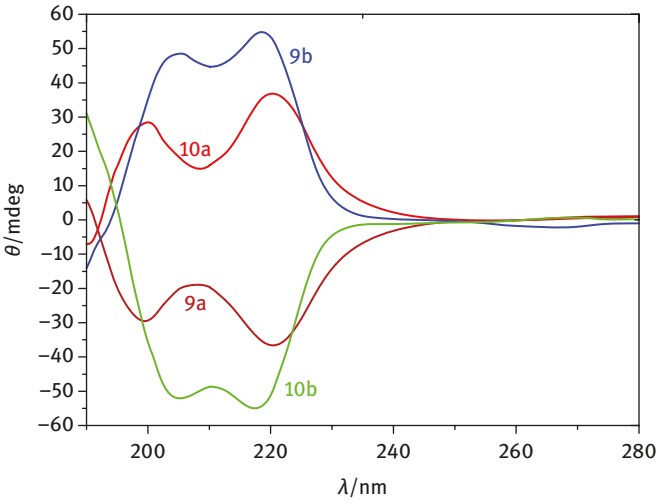


Figure 6.13: Solid-state CD spectra of compounds (9a–10b).

Some principles are summarized after the systematic study and the absolute configuration of central phosphorus atom may dominate the Cotton effect for this kind of compounds.

The chelated amino acids of compounds (**3a–4b** and **5a–6b**) are valine and leucine, whose side chains are saturated fatty chains. So the Cotton effect of solid CD spectra is affected strongly by the absolute configuration of central phosphorus, but not the α-carbon of amino acid. When the absolute configuration of central phosphorus is Δ_P, resulting in the positive Cotton effect and vice versa.

When the side chain of amino acid is changed to aromatic ring (compounds **7a–10b)** as a strong chromophore in solid CD spectra, the Cotton effect could be impacted collectively by both α-carbon of amino acid and absolute configuration of central phosphorus. So their solid CD spectra show a special form.

The phenyl ring of compounds **7a–8b** is closer to the α-carbon of amino acid than that of compounds **9a–10b**, so its Cotton effect is deeply impacted by this chromophore. And the Cotton effect of compounds **9a–10b** is dominated by absolute configuration of central phosphorus, which is in agreement with that of compounds **3a**–**6b.**

In conclusion, 16 chiral pentacoordinate spirophosphoranes with bis-α-amino acid bonds were synthesized effectively and their stereochemistry was studied systematically by several methods such as NMR, solid CD and X-ray spectra. All enantiomers of our compounds share the mirror-image symmetric CD spectra and uniform ^{1}H NMR. Otherwise, the diastereoisomer could be distinguished by some parameters such as the coupling constants $^1J_{P\text{-}X}$ and $^4J_{H\text{-}C\text{-}N\text{-}P\text{-}H}$. By the theoretical calculation of quantum chemistry, the spectra of ^{1}H NMR were reasonably explained. Since phosphoryl transfer processes in biosystems are highly regioselective and

substrate specific, the present results suggest that pentacoordinate phosphorus is a very important structural feature relevant to chiral phosphoryl transfer pathways.

References

[1] 0Ji G J, Xue C B, Zeng J N et al. Synthesis of N-(Diisopropyloxyphosphoryl) amino acids and peptides [J]. *Synthesis* 1988*;* 1988(6): 444–448.

[2] Li Y M, Yin Y W, Zhao Y F. Phosphoryl group participation leads to peptide formation from N-phosphorylamino acids [J]. *International Journal of Peptide and Protein Research* 1992**,** *39* (4), 375–381.

[3] Zhao Y, Ju Y, Li Y et al. Self-activation of N-phosphoamino acids and N-phosphodipeptides in oligopeptide formation [J]. *International Journal of Peptide and Protein Research* 1995**,** *45* (6), 514–518.

[4] Fu H, Li Z L, Zhao Y F, et al. Oligomerization of N,O-bis(trimethylsilyl)-alpha-amino acids into peptides mediated by o-phenylene phosphorochloridate [J]. *Journal of the American Chemical Society* 1999**,** *121* (2), 291–295.

[5] Lin C X, Li Y M, Cheng C M et al. Penta-coordinate phosphorous compounds and biochemistry [J]. *Science China Series B* 2002**,** *45* (4), 337–348.

[6] Li Y M, Zhao Y F. The bioorganic chemical reactions of N-phosphoamino acids without side chain functional group participated by phosphoryl group [J]. *Phosphorus, Sulfur, and Silicon and the Related Elements* 1993**,** *78* (1), 15–21.

[7] Li Y C, Tan B, Zhao Y F. Phosphoryl transfer reaction of phospho-histidine [J]. *Heteroatom Chemistry* 1993**,** *4* (4), 415–419.

[8] Xue C B, Yin Y W, Zhao Y F. Studies on phosphoserine and phosphothreonine derivatives: N-diisopropyloxyphosphoryl-serine and -threonine in alcoholic media [J]. *Tetrahedron Letters* 1988**,** *29* (10), 1145–1148.

[9] Westheimer F H. Pseudo-rotation in the hydrolysis of phosphate esters [J]. *Accounts of Chemical Research* 1968**,** *1* (3), 70–78.

[10] Ramirez F. Oxyphosphoranes [J]. *Accounts of Chemical Research* 1968**,** *1* (6), 168–174.

[11] Swamy K CK, Kumar N S. New features in pentacoordinate phosphorus chemistry [J]. *Accounts of Chemical Research* 2006**,** *39* (5), 324–333.

[12] Gololobov Y G, Gusar N I, Chaus M P. Reactions of trivalent phosphorus compounds with azides containing a mobile H-atom: A conception of phosphazo-compound spirocyclization mechanism [J]. *Tetrahedron* 1985**,** *41* (4), 793–799.

[13] Zaloom J, Calandra M, Roberts D C. A new synthesis of peptides from azides and unactivated carboxylic acids [J]. *The Journal of Organic Chemistry* 1985**,** *50* (14), 2601–2603.

[14] Abd-Ellah I M, Ibrahim E H M, El-khazander A N. Studies in cyclodiphosphazanes: Some reactions of hexachlorocyclodi-phosphazames(I) with amino acids [J]. *Phosphorus and Sulfur and the Related Elements* 1987**,** *31* (1), 13–18.

[15] Garrigues B, Munoz A, Koenig M et al. Spirophosphoranylation d'[alpha]-aminoacides--I: Preparation et tautomerie [J]. *Tetrahedron* 1977**,** *33* (6), 635–643.

[16] Munoz A, Garrigues B, Wolf R. Preparation de composes du phosphore tri-et pentacoordinine a partir D'hydroxyacides et D'aminoacides [J]. *Phosphorus and Sulfur and the Related Elements* 1978**,** *4* (1), 47–52.

[17] Garrigues B, Munoz A, Mulliez M. Hydrolyse et alcoolyse de spirophosphoranes contenant le ligand α-aminoacide [J]. *Phosphorus and Sulfur and the Related Elements* 1980**,** *9* (2), 183–188.

[18] Hou J B, Zhang H, Guo J N et al. Chirality at phosphorus in pentacoordinate spirophosphoranes: Stereochemistry by X-ray structure and spectroscopic analysis [J]. *Organic & Biomolecular Chemistry* 2009, *7* (15), 3020–3023.
[19] Gorenstein DG. Phosphorus-31 NMR Principles and Applications[M]. Orlando, Florida: Academic Press. Inc. 1984.
[20] Hou J B, Tang G, Guo J N et al. Stereochemistry of chiral pentacoordinate spirophosphoranes correlated with solid-state circular dichroism and H-1 NMR spectroscopy [J]. *Tetrahedron-Asymmetry* 2009, *20* (11), 1301–1307.

7 A new theoretical model for the origin of amino acid homochirality

Amino acid homochirality of life is a major scientific topic. Why are proteins made up of L-enantiomer, whereas nucleic acid polymers such as DNA and RNA associated with D-sugars? In this chapter, phosphoryl amino acid-5′-nucleosides with P–N bond are postulated to be a chiral origin model in prebiotic molecular evolution. Based on the theoretical model, our experiments and calculations show that the chiral selection of the earliest amino acids for L-enantiomers seems to be a strict stereochemical/physicochemical determinism. As other amino acids developed biosynthetically from the earliest amino acids, we infer that the chirality of the later amino acids was inherited by the precursor amino acids. This idea probably goes far back in history, but it is hoped that it will be a guide for further experiments in this area.

7.1 The origin of homochirality

The origins of genetic code and homochirality are the two most important topics in research with origin of life. Genetic code is an algorithm that relates each amino acid to one or more trinucleotide sequences. The genetic code-loaded L-shaped tRNA, which came into existence around 1 billion years ago, has remained basically unchanged over the ages and has been adopted by almost all living organisms [1]. The stereochemical hypothesis suggests that the origin of the genetic code can probably be attributed to stereochemical and/or physicochemical interactions between anticodons or codons and amino acids [2], whereas the coevolution hypothesis argues that the biosynthetic pathways of amino acids played a fundamental role in defining the organization of the genetic code [3].

Homochirality means that the monomer units of proteins are made up of the L-enantiomer, whereas the monomer units of the nucleic acid polymers such as DNA and RNA as well as those of the biologically important polysaccharides are associated with D-sugars [4]. Therefore, it is now recognized that all of the crucial biopolymers associated with life are homochiral. The reason behind the homochirality is still a puzzle and several controversial hypotheses have been presented. It might have stemmed from an asymmetric adsorption on chiral mineral surfaces such as quartz or clays, from the consequences at the molecular level of the violation of parity in the weak interaction (the so-called Vester–Ulbricht hypothesis), from a spontaneous resolution or from asymmetric photoreactions [5, 6]. However, there are few reports on the study of the relation between the origin of genetic code and amino acid homochirality. We believe that at the stages of the origin of the genetic code and amino acid homochirality, their purely chemical evolution processes not only dominated but also cooperated with evolution. The most archaic forms of genetic code established

https://doi.org/10.1515/9783110562552-007

the physicochemical interactions between nucleotides and amino acids, whose constraints of steric interaction might affect the chiral selection of amino acids. Therefore, we suggest that the amino acid homochirality, as a unique behavior of life, might have originated synchronously with the genetic code; in other words, this specific interaction between amino acids and short oligonucleotides could have produced a chiral selectivity of D- and L-enantiomers, which consequently broke the chiral in the origin of the genetic code.

In this study, it is proposed that the formation of phosphoryl amino acid-5′-nucleosides with P–N bond could have been important as the model of origin in amino acid chiral selectivity. Its analogous compounds such as aminoacyl-3′, 5′-nucleotides [7] have once been proposed as a precursor to a close coding of amino acid side chains and nucleic acid bases. Here, the model of the phosphoryl amino acid-5′-nucleosides is justified as arising from a long and systematic study of the function of phosphoryl amino acids and phosphoryl nucleosides. We have found that the P–N bond in phosphoryl amino acids could be transferred to peptide bonds and phosphodiester bonds through reaction with amino acids and nucleosides [8]. And then the reactions of *N*-phosphoryl amino acids such as phosphoryl amino acid-5′-nucleosides are considered as a model to study many processes of origin of life, such as chiral origin and genetic code origin. The model presented here is clearly speculative, but it is hoped that it will serve as a guide for further experiments.

7.2 Model for the origin of amino acid homochirality

The spontaneous interaction between free amino acids and ribonucleotides occurs at a very low rate in water system. There is no doubt that clays, sands and many minerals can concentrate biomolecules and catalyze their polymerization [9]. Among these substances, since the minerals with metal ions have a stronger catalyzing capability, the surface adsorptions of minerals are much more important. These positively charged minerals are of interest for negatively charged molecules for selective concentration, ordering and surface-induced polymerizing reactions. If biomolecules are to be negatively charged, first consideration should be given to phosphate as the molecules with phosphates play an essential role in the origin of life. For example, the produce of the phosphoribose backbone, such as DNA and RNA, structurally depends on the characteristics of phosphates as the linking group, and amino acid phosphates may serve as an activation precursor to peptide chemistry [10]. Where did phosphates come from and how could the inorganic phosphorus perhaps have been incorporated into organic molecules? Volcanic areas are known to be an important cradle land of life. By volcanic activity, lots of phosphatic rocks gushed up to the ground. From both simulation experiments and analysis of volatile condensates in volcanic gas, the phosphatic rocks can produce water-soluble polyphosphates consisting of

PO_4^{3-}, $P_3O_9^{3-}$, linear oligophosphate and cyclic triphosphate (P_{3m}) through partial hydrolysis of P_4O_{10}. Under sterile conditions at Mg^{2+}/Ca^{2+} ratios of natural waters, most phosphates from hydrolysis of polyphosphates would precipitate as insoluble salts [11]. Whereas P_{3m}, a six-membered cyclic phosphate, readily undergoes a ring-opening reaction in the presence of amino acids and nucleosides to give Pi-(*N*) amino acid with a P–N bond and mononucleotide [12]. And then gradually gives rise to the phosphoryl amino acid-5′-nucleosides (Figure 7.1). Figure 7.1 shows that the phosphate plays a crucial role in earlier stages of the origin of life: the charged phosphate groups not only serve as surface-bonding mediation (i.e., surface affinity) but also fulfill the functions of connector and activation.

Considering the conjecture that all kinds of polymers could be produced in phosphoryl system in Figure 7.1, we carried out a preliminary study in our previous experiments [13]. For example, the reaction products of *N*-(*O*,*O*-diisopropyl) phosphothreonine (DIPPThr) and four nucleosides (adenosine (A), uridine (U), cytidine (C) and guanosine (G)) at room temperature in anhydrous pyridine were reported as follows: $(Thr)_{2\text{-}4}$, $DIPP(Thr)_{2\text{-}4}$, XpX, $(Xp)_2$, $(Xp)_2$-Thr and so on were determined by fast atom bombardment mass spectrometry (FAB-MS), High Performance Liquid Chromatography (HPLC), ultraviolet (UV) and capillary electrophoresis-mass spectrometry (CE-MS) techniques (X represents A, U, C and G). The occurrence of the above polymers suggests that the peptides and oligonucleotides may have been formed simultaneously on

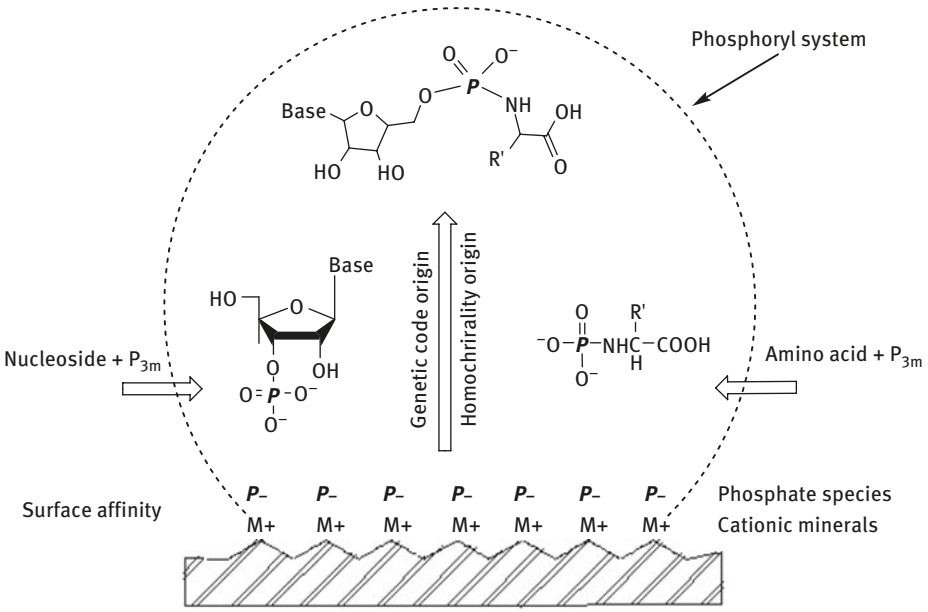


Figure 7.1: The possible functions of phosphates in molecular evolution.

the primitive earth. In particular, the occurrence of the $(Xp)_2$-Thr product and the strong hydrolytic stability of phosphodiester bonds [14] have led us to speculate that the free amino acids and ribonucleotides could have been connected by the charged phosphodiester bonds during chemical evolution. The formation of the ester linkage was extremely favorable for the interaction between amino acids and ribonucleotides. Thus, the specific amino acid and nucleotide conjugated compound could be identified as the progenitor of the present genetic code and amino acid homochirality. Based on the above hypothesis, a structural model could be build (as shown in Figure 7.2). According to the code triplet theory [15], the first two nucleotides generate a stereospecific interaction with the amino acid side chain, with the second nucleotide being vital. For this reason, the phosphate is esterified at the 5′ position in our model, which means that it is connected with the second nucleotide of the code triplet in the structure shown in Figure 7.2. The above hypothesis goes far back in history; nevertheless, some primary deductions have been confirmed by our modern experiments and theoretical calculations.

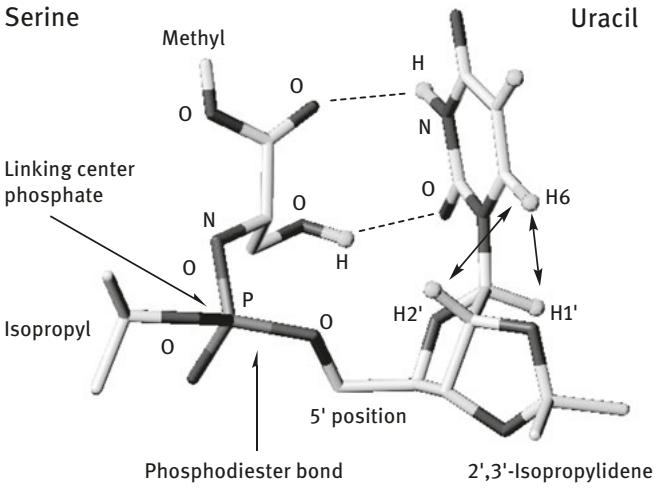


Figure 7.2: The structure model of the amino acid-5′-nucleotides with P–N bond, similar to the Ser:Urd compound (dashed line stands for H-bond).

7.3 Interaction of nucleosides and amino acids

In an effort to ascertain the steric relationship between nucleoside bases and amino acid side chains, we synthesized a series of nucleoside derivatives, such as 2′, 3′-isopropylideneuridine 5′-isopropyl methoxyserinyl phosphamide (Ser:Urd compound). The general synthetic procedure is illustrated in Figure 7.3. The conformation of the synthesized compounds is characterized by the nuclear Overhauser effect (NOE) of nuclear magnetic resonance technique. As shown in Figure 7.4, the uracil's H_6 gives notable NOE signals with respect to H_1' and H_2' of the ribose ring, indicating that the Ser:Urd compound is in favor of the *Syn* conformation. This can

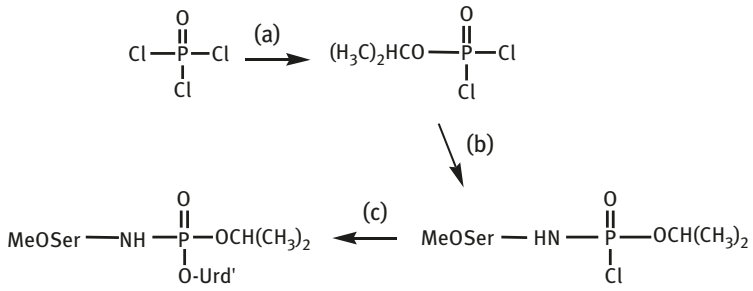


Figure 7.3: Reagents and condition of the Ser:Urd compound: (a) isopropanol, triethylamine, anhydrous diethyl ether, 0°C , 2 h , then rt , 5 h; (b) L-serine methyl ester hydrochloride, triethylamine, anhydrous CH_2Cl_2, –70°C , 2 h, then rt , 6 h; (c) 2′,3′-isopropylideneuridine (Urd′), triethylamine, anhydrous THF, rt, 24–48 h.

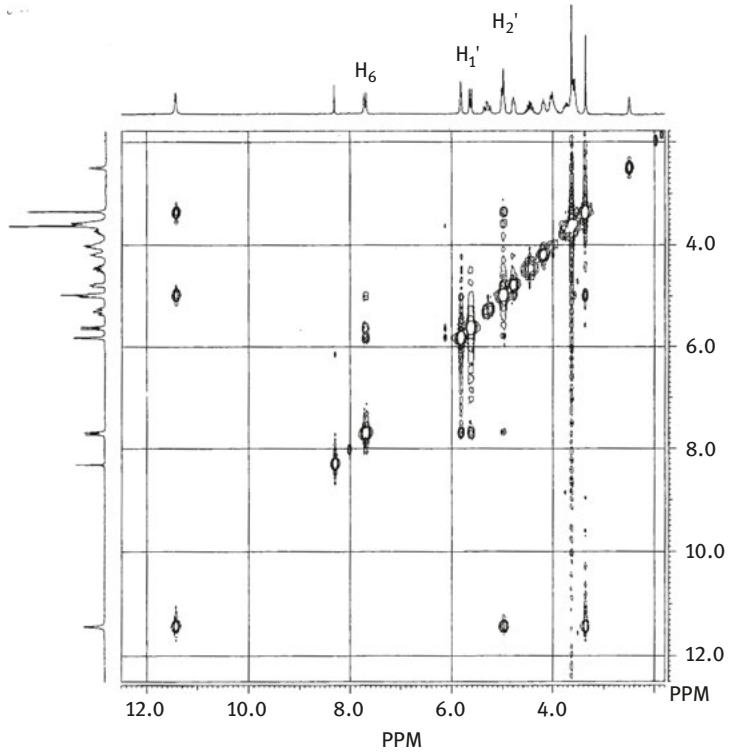


Figure 7.4: The 500 MHz NOE spectra of the Ser:Urd compound in DMSO.

be attributed to the two pairs of hydrogen bonds that have formed between carbonyl and hydroxyl groups of serine and the uridine base, with the first pair between the O atom of serine carbonyl group and the base H (N_3) and the second pair between the

base O_2 and the H atom of the serine β-OH (Figure 7.2). The above results show that the conformation of the Ser:Urd compound is decided by the strong interaction of the hydroxyl and ester groups of serine with uracil base. Similarly, the steric interactions between nucleoside bases and amino acid side chains also exist in the other phosphoryl amino acid-5′-nucleosides. For nonpolar side chains such as Ala and Val, hydrogen bonds could be formed between carboxyl groups of amino acids and bases, whereas aromatic amino acids allow significant aromatic–aromatic overlap with the bases. This has provided an extremely important basis for experiments with our theoretical model of chirality recognition.

7.4 Thermodynamic parameters of interaction between nucleosides and amino acids

In order to identify which chiral selection would occur in terms of the structural model in Figure 7.2, we calculated the formation potentials of about 160 phosphoryl amino acid-5′-nucleosides (20 amino acids × 4 ribonucleosides × 2 enantiomers). As the products of all amino acid:ribonucleotide reactions have the same bond breakage or formation, the relative formation potential simply equals the difference between the total potential of isolated fragments (such as amino acid, ribonucleoside and phosphate) and the potential of the compound system [16]. Therefore, the relative formation potential (i.e., the potential difference after and before reaction) aptly exhibits the stability difference of the conformations with the specific steric interactions between the chiral amino acid and ribonucleotide. First, on the basis of Figure 7.2, a starting structure of phosphoryl amino acid-5′-nucleosides was manually built by making amino acid side chains as close to nucleoside bases as possible to form H-bonds; the compounds were then hydrated by a 2 nm box of TIP3P water molecules. Afterward, the optimum conformation of each compound was determined by conformational sampling of molecular dynamics (MD). The MD protocol contained an equilibration part lasting 10 ps and a sample part lasting 1 ps in water box at 298 K. Each 0.2 ps was sampled for a conformation, and then the five conformations of the compound extracted from the water box were minimized. Among them, the potential of the optimum conformation was used for the calculation of the relative formation potential. The fragment potentials were calculated by energy minimization with a conjugate gradient optimizer at a convergence criterion of 4.18 kJ/mol/nm. All calculations were performed on a 3.0G PC utilizing MM+ force field in HyperChem 6.0 consistently using default settings (assigning a dielectric constant of 1.0).

The relative formation potentials from random associations between chiral amino acids and ribonucleotides are shown in Table 7.1 and Figure 7.5. The bold line represents the formation potentials of L-amino acid compounds (L-aas) while the dashed line stands for D-amino acid compounds (D-aas). With lower formation potentials, the

Table 7.1: The relative formation potentials of 156 amino acid-5′-nucleotides (kJ/mol)[a]

	Adenosine		Cytidine		Guanosine		Uridine	
	L	D	L	D	L	D	L	D
Gly	*17.05*		22.95		19.35		25.75	
Ala	31.64	35.03	34.53	34.44	*29.89*	31.31	33.23	31.89
Ser	36.45	45.94	31.10	32.06	*23.49*	27.50	31.39	33.02
Asp	29.80	36.37	33.48	37.79	*22.91*	30.22	35.40	39.04
Glu	*39.21*	47.36	44.27	48.95	41.34	45.48	43.72	49.95
Val	29.68	39.21	33.44	40.55	*24.12*	37.54	35.82	42.68
Pro	45.06	54.17	*38.96*	40.67	43.18	43.35	45.60	55.18
Thr	23.66	*16.93*	25.71	19.77	24.79	19.14	27.63	20.98
Leu	41.63	41.17	43.10	39.79	*26.38*	38.87	48.15	53.46
Ile	23.41	37.95	27.34	31.39	*20.27*	22.70	29.76	33.36
Phe	31.81	25.92	44.60	38.58	32.81	*23.95*	45.48	39.17
Tyr	26.54	16.47	38.50	24.62	21.15	*13.13*	35.99	16.55
Cys	37.54	26.67	29.55	*25.21*	30.76	26.71	32.06	26.88
Lys	19.23	26.42	*17.39*	31.10	17.60	22.86	19.27	22.95
Arg	35.36	35.99	26.88	*20.90*	28.88	31.18	30.68	25.08
His	34.86	*17.35*	40.09	30.10	30.68	19.98	40.34	32.56
Met	27.42	34.32	*11.66*	22.15	30.01	31.22	36.07	39.38
Trp	28.67	28.42	36.28	35.20	27.63	*25.71*	39.08	37.66
Asn	33.86	43.26	23.37	43.81	*18.98*	42.76	37.62	34.40
Gln	33.02	46.11	20.06	30.93	*11.45*	15.72	20.77	28.80

[a]Values in italics represent the minimum potential for each amino acid, respectively.

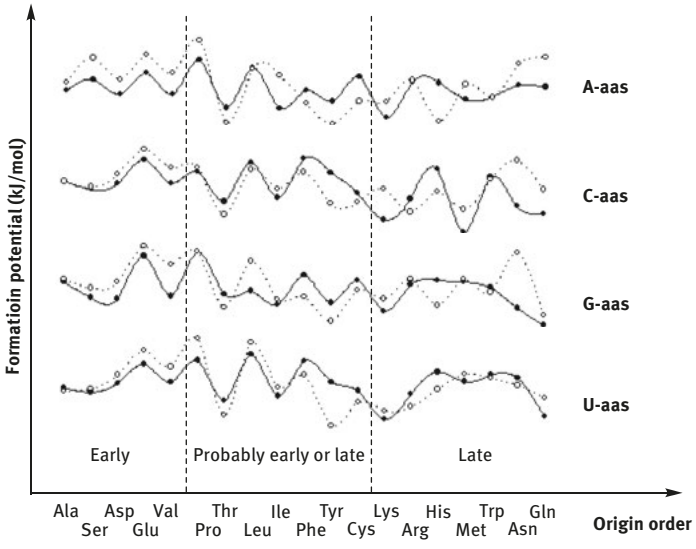


Figure 7.5: The relative formation potentials of amino acid-5′-nucleotides (the bold line represents L-aas; the dashed line stands for D-aas).

amino acids would be more apt to bind with ribonucleotides. Statistics data indicate that the rate of the L-amino acids being selected by ribonucleotides for binding is 59.21% and that of the D-amino acids is 40.79%. Corresponding to Figure 7.5, the two linking lines of the formation potential appear almost crossed. The sequence of the amino acids from left to right follows their original order in Figure 7.5. Among them, Gly (without chirality), Ala, Ser, Asp, Glu and Val are the earliest amino acids [17]. The rate of the earliest L-amino acids being selected stands at 90.00% and that of the D-amino acids stands at 10.00%. Clearly, bold lines are lower than dashed lines on the left of Figure 7.5; in other words, as far as the earliest amino acids are concerned, the rate of L-amino acids being selected is remarkably larger than that of D-amino acids. The formation potential is an evaluating parameter of the conventional thermodynamic approaches. Here, based on the modeling method, the formation potential mainly consists of hydrogen bonding, van de Waals, electrostatic and so on. These interactions notably depend upon steric constraints and orientation between amino acid side chains and the bases of ribonucleotides, and their physicochemical properties. Therefore, the chiral selection of the earliest amino acids seems to be a strict stereochemical /physicochemical determinism, which is highly consistent with the origin of the genetic code [18].

For some other amino acids, especially the later amino acids, based on the relative formation potentials, L-enantiomers are not more prior to bind with ribonucleotides than D-enantiomers (the cross section of the two lines in Figure 7.5). This indicates that the chiral selection of the later amino acids has somehow abandoned the strict stereochemical/physicochemical determinism; in other words, stereochemical/physicochemical hypothesis alone is not sufficient to account for the chiral selection of the later amino acids. Similar to the coevolution theory of the genetic code origin, we infer that the amino acids that had first appeared (the precursors), possibly five (except for Gly), were selected L-enantiomers. As other amino acids developed biosynthetically from these precursors, the chirality of the precursor amino acid in life was inherited by the product amino acids. Finally, with the evolution of amino acids, the homochirality in life gradually came into completion. For the difference in formation potentials of L, D-enantiomers between the earliest and later amino acids, it can be explained as follows: As the earliest amino acids used to have the simpler side chain, the asymmetry environment provided by the five-membered ring of D-sugar has a notable effect on the conformation potentials. As the structure of side chains becomes more complicated, the D-sugar asymmetry affection gradually decreases. Therefore, we infer that the chirality and asymmetry of sugar could determine the chiral selection of amino acids.

Additionally, the modeling results show that besides the chirality, amino acids also have selection for ribonucleotides through the phosphodiester conjugation. For the major amino acids, their minimum formation potentials are concentrated in guanine and cytosine regions, especially in the guanine region (italic values in Table 7.1). The code relationships are weak in the case of the model of origin of amino

acids versus ribonucleotides (1:1). However, it should be noted that this seems to provide the simplest model of the genetic code. At the very start of the genetic code evolution, the original direct matching between short oligonucleotides and amino acids might exist. Moreover, the codon–amino acid assignment was hypothesized to evolve from simplicity to complexity: the amino acids were only codified by cytosine and guanine in the early genetic code, and then expanded from the G and C toward the complete G, C, A and U [19]. Here, the formation potentials of G and C compounds are lower, suggesting that the selection of the G and C ribonucleotides as the precursor codons might be related to their own chemical properties.

Through the application of molecular modeling, this work has characterized a new theoretical model of the origin of amino acid homochirality, with some meaningful results obtained. First, we suggest that in the earlier stages of the origin of chirality, the steric interactions between amino acid side chains and nucleotide bases could promote a constant and strong selective drive of L-enantiomers, and that subsequently, the chirality in life was inherited with the evolutionary development of the amino acids themselves. Second, based on the theoretical model of nucleotides versus amino acids (1:1), the binding system can not only be quite selective for earlier L-amino acids over the D-enantiomer, but also distinguish the bases of nucleotides, especially for cytosine and guanine. The work described in this chapter is very helpful in better understanding the relationship between amino acid: nucleotide affinity and the genetic code.

References

[1] Gargaud M, López-Garcìa P, Martin H. Origins and Evolution of Life: An Astrobiological Perspective[M]. Cambridge University Press, Cambridge, England, 2011.

[2] Di Giulio M. Physicochemical optimization in the genetic code origin as the number of codified amino acids increases [J]. *J Mol Evol*, 1999, 49:1–10.

[3] Di Giulio M. Reflections on the origin of the genetic code: A hypothesis [J]. *J Theor Biol*, 1998, 191: 191–196.

[4] Bonner W A. Parity violation and the evolution of biomolecular homochirality [J]. *Chirality*, 2000, 12: 114–126.

[5] Plasson R, Tsuji M, Asakura K. Reactivity of alanylalanine diastereoisomers in neutral and acid aqueous solutions: a versatile stereoselectivity [J]. *Origins Life Evol Biosph*, 2011, 41: 413–435.

[6] Apontea J, Tarozoa R, Alexandreb M et al. Chirality of meteoritic free and IOM-derived monocarboxylic acids and implications for prebiotic organic synthesis [J]. *Geochim Cosmochim Ac*, 2014, 131: 1–12.

[7] Han D X, Wang H Y, Ji Z L et al. Amino acid homochirality may be linked to the origin of phosphate-based life [J]. *J Mol Evol*, 2010, 70: 572–582.

[8] Han D X, Chen W, Han B, Zhao Y F, A new theoretical model for the origin of amino acid homochirality [J]. *Sci China Ser C*, 2007, 37: 382–388.

[9] Fuchida S, Masuda H, Shinoda K. Peptide formation mechanism on montmorillonite under thermal conditions[J]. *Origins Life Evol Biosph*, 2014, 44: 13–28.

[10] Baltscheffsky H, Blomberg C, Liljenstrom H et al. On the origin and evolution of life: an introduction [J]. *J Theor Biol*, 1997, 187: 453–459.
[11] Kolb V, Zhang S B, Xu Y et al. Mineral induced phosphorylation of glycolate ion – a metaphor in chemical evolution [J]. *Origins Life Evol Biosph*, 1997, 27: 485–503.
[12] Tsuhako M, Fujimoto M, Ohashi S et al., Phosphorylation of nucleosides with cyclo-triphosphate[J]. *Bull Chem Soc Jpn*, 1984, 57: 3274–3280.
[13] Zhou W H, Ju Y, Zhao Y F et al., Simultaneous formation of peptides and nucleotides from N-phosphothreonine [J]. *Origins Life Evol Biosph*, 1996, 26: 547–560.
[14] Arrhenius G, Sales B, Mojzsis S et al., Entropy and charge in molecular evolution-the case of phosphate[J]. *J Theor Biol*, 1997, 187: 503–522.
[15] Seligmann H, Amzallag G N. Chemical interactions between amino acid and RNA: multiplicity of the levels of specificity explains origin of the genetic code [J]. *Naturwissenschaften*, 2002, 89: 542–551.
[16] Yang P, Han D X. Molecular modeling of the binding mode of chiral metal complexes $[Co(phen)_2dppz]^{3+}$ with B-DNA[J]. *Sci China B*, 2000, 43: 516–523.
[17] Di Giulio M. The coevolution theory of the origin of the genetic code. *Phys Life Rev* [J], 2004, 1: 128–137.
[18] Di Giulio M, Medugno M. The historical factor: the biosynthetic relationships between amino acids and their physicochemical properties in the origin of the genetic code [J]. *J Mol Evol*, 1998, 46: 615–621.
[19] Ikehara K, Omori Y, Arai R et al. A novel theory on the origin of the genetic code: a GNC-SNS hypothesis [J]. *J Mol Evol*, 2002, 54: 530–538.

8 *N*-Phosphoryl amino acids and the origin of cell membranes

The formation and evolution of the basic structure of cells are some kinds of significant biological evolution events. Therefore, the study on the origin of the cell itself is an important part of exploring the origins of life and evolution of biospheres. Due to the particular nature of *N*-phosphoryl amino acids, it could trigger the formation of oligonucleotides, oligopeptides and phospholipids simultaneously to realize the "three in one" origin of the prototype cells.

8.1 The origin and evolution of cells

8.1.1 The formation of primitive cells – The beginning of the music of life

The original Earth was formed about 4.6 billion years ago. According to the famous "primordial soup" hypothesis, in the early stage of Earth formation, inorganic small molecules gradually accumulated to produce organic small molecules in the turbulent environment. In addition, there is also a possibility that the organic small molecules that formed in the universe arrived on the Earth through the meteorite impact and other forms. These conditions provided sufficient nutrients for the appearance of primitive life. In the endless chemical evolution process of the original Earth, these nutrients gradually formed the precursor materials of modern biological macromolecules, such as amino acids and nucleic acid bases. Soon afterward, these building blocks polymerized into the complex biological macromolecules, and gradually produced large macromolecular complex in the primitive ocean, that is, proteinoid. Along with the emergence of genetic code (nucleic acid), the primitive cells with the self-reproduction function are formed.

Cells are the basic structural, functional units of all the living organisms. However, the primitive cells are very fragile. Up to now, they have become extinct along with the biological evolution. Since it is extremely hard to keep their related geological records, their structural features are still unclear. According to the current research results, we could conclude that the primitive cells do not have any organelles except ribosome, but they already have independent cell membranes and RNA with the most basic physiological functions. Besides, the primitive cell can also complete the replications of genetic materials and cell membrane to achieve cell proliferation process. Therefore, the core issue of the origin of primitive cell is precisely the origin of the cell membrane. The emergence of cell membrane is an important symbol of the origin of life, as well as an important milestone in the transformation process of the former life material into the modern life.

https://doi.org/10.1515/9783110562552-008

8.1.2 The formation of prokaryotic and eukaryotic cell

According to the records on fossils, prokaryocyte first appeared on the Earth about 3.0–3.5 billion years ago. Their main features are that they lack membrane-bound nucleus, just nucleoid, and have low evolutionary status. However, prokaryotic cells are already a class of very well-structured cells. It is hard to imagine how they evolved from noncell life forms to primitive cells and then evolved into prokaryotic cells in harsh primitive earth environments. At this point, DNA and protein have been gradually ascended to the stage of history, gradually replacing the partial functions of RNA in primitive cells, which is another important sign of biological evolution.

A large number of facts based on molecular biology and ancient microbiology have shown that prokaryotic cells and eukaryotic cells have a common origin. Moreover, prokaryotic cells appear earlier than eukaryotic cells in the history of biological evolution. From the point of cell structures, the most critical step is the formation of nucleus in the evolution from prokaryotic cells to eukaryotic cells. The nucleus is the most important organelle in eukaryotic cells, which composed mainly of nuclear membrane, chromatin, nucleus and nucleoplasm. The most critical point of the origin of the nucleus is the origin of the nuclear membrane, because the protocells have no nuclear membrane. At present, there are three main views on the origin of nuclear membrane: First is that the nuclear membrane is produced by the inner fold of the cell membrane, enclosing the original nucleus; the other is that the nuclear membrane of the inner and outer membranes has different origins, the inner membrane is from the cell membrane, while the outer membrane is derived from endoplasmic reticulum; there is also a view that the nuclear membrane originated from the original endoplasmic reticulum formed by the cell membrane. No matter what kind of theory, it has shown that nuclear membrane and cell membrane are inextricably linked.

Therefore, the study of the origin of the cell membrane and its chemical properties is one of the important ways to understand the origin and evolution of life, and to explain the law of life activities.

8.2 Membrane structures constructed by amphiphilic *N*-phosphoryl amino acids

It is well known that the function of cell membranes in life is less important than DNA and protein, except viruses. Why is it hardly to find the life forms without cell in nature? This is because cell membranes provide a relatively stable living environment for life. In 1958, B. Davis published an article entitled “The Importance of Ionization” [1], clarifying that the cell membrane must be ionized to effectively keep life substances in the cell. Otherwise, the neutral molecules can be miscible with the cell membrane, resulting in the transmission of neutral molecules from the cell

membrane, and the infinite dilution. Therefore, for ionizable cell membranes, it is necessary to simultaneously bridge hydrophilic groups and hydrophobic long chains and satisfy its own charge characteristics, and it must be at least trivalent. At this point, the phosphoric acid is the most likely to be thought of. And for other bridging atoms, valence and the instability of the compound make its biological activity much weaker than phosphoric acid, showing that the phosphate group is very important as an indispensable structure in the cell membranes.

In the early stage of the origin of life, the self-assembly of amphiphilic molecules played a very important role [2, 3] during the process of the protocell formation. This kind of ordered assembly can effectively provide a template for the initial synthesis of biological macromolecules. A plenty of experiments show that taking ordered layered assemblies formed by phospholipid, or clay and other mineral surfaces as templates [4–6], or making the amphiphilic derivatization of amino acids to form an ordered layered composite can promote the reaction of amino acid condensation into peptides.

N-Phosphoryl amino acids, a mini-activated enzyme, were discovered by our research group in the earlier study. According to the structural characteristics of phospholipids, the amphiphilic derivatization of *N*-phosphoryl amino acids was carried out and the amphiphilic long-chain alkoxy-*N*-phosphoryl amino acid molecular model [7] was proposed.

Comparing the chemical structure of long-chain alkoxyl *N*-phosphoryl amino acids with the phospholipids of the classic biofilm, it can be found that the two structures are very similar in structure: both contain hydrophilic polar head groups and lipophilic nonpolar alkyl long chains (as shown in Figure 8.1). Thus, amphiphilic *N*-phosphoryl amino acids have the potential to become the original cell membrane. Moreover, *N*-phosphoryl amino acids can be embedded in membrane phospholipid bilayers, regulating the fluidity of phospholipids [8]. Therefore, the model molecule has important significance in biofilm simulation, and it has an active effect on explaining the origin of *N*-phosphoryl amino acid in cell membrane.

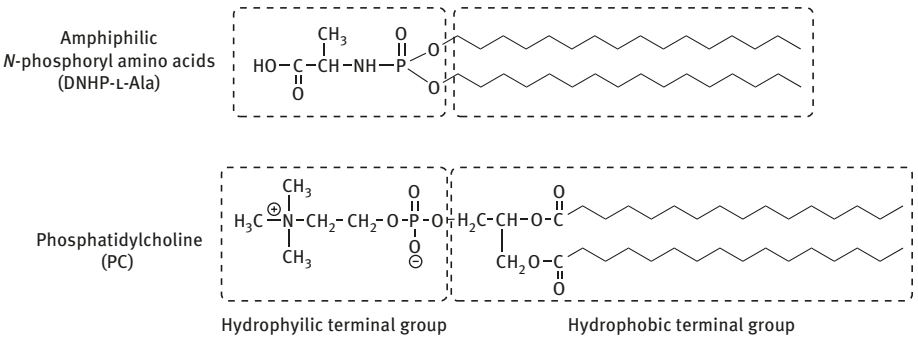


Figure 8.1: Structures of long-chain alkoxy *N*-phosphoryl amino acids and biofilm phospholipids.

8.2.1 Synthesis of amphiphilic *N*-phosphoryl amino acids

The compound capable of forming the film is usually amphiphilic. In general, it contains both a hydrophilic head group and a hydrophobic tail chain. In our laboratory, we designed and synthesized a series of amphiphilic *N*-phosphoryl amino acids with various lengths of hydrophobic tails and the hydrophilic alanine head groups through the improved Todd method. The detailed synthetic methods are shown in Figure 8.2, and the spectral characterization data for the relevant compounds are summarized in Table 8.1.

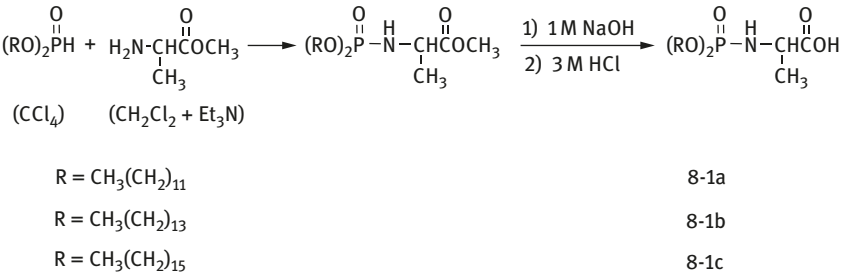


Figure 8.2: Synthesis of long-chain alkoxyl *N*-phosphoryl amino acids.

8.2.2 Self-assembled monolayers of amphiphilic *N*-phosphoryl amino acids at water–air interface and their condensation reactions

8.2.2.1 Formation of monomolecular membranes from amphiphilic *N*-phosphoryl amino acids at the water–air interface

The surface pressure–single molecule area isotherm, that is Π–*A* isotherm, is one of the most important properties of monomolecular films. It can be used to characterize the relationship between the two-dimensional space pressure and the concentration of monomolecular films and reflect the spatial arrangement and change of the amphiphilic molecules in monomolecular films. The aggregation behavior of the amphiphilic molecules at the water–air interface, that is, the quality and properties of the monomolecular films depend not only on the properties of the amphiphilic molecules (such as the molecular sizes and structures), but also on the subphase environment (such as temperature and pH). In our previous research works, the effects of the amphiphilic *N*-phosphoryl amino acid molecular structures and the subphase environment on the properties of monomolecular films were investigated through Π–*A* isotherm.

The results showed that the amphiphilic *N*-phosphoryl amino acid can form a stable monomolecular membrane at the water–air interface. The structures of film-forming molecules, the headgroup conformation, the subphase environment

Table 8.1: Spectra data of long-chain alkoxyl ***N***-phosphoryl alanine.

Compounds	Physical State (room temperature)	ESI-MS (m/z, M + H^+)	^{31}P-NMR (δ, ppm)	^{1}H-NMR (δ, ppm)	IR (cm^{-1})
8-1a	Colorless liquid	506.2	9.43	0.876 (t, 6H, $C\underline{H}_3(CH_2)_{11}O$–), 1.255 (m, 36H, $CH_3(C\underline{H}_2)_9CH_2CH_2O$–), 1.437 (d, 3H, –CH($C\underline{H}_3$)COOH), 1.625 (m, 4H, $C_{10}H_{21}C\underline{H}_2CH_2O$–), 3.816 (m, 1H, –C$\underline{H}$($CH_3$)COOH), 3.969 (m, 4H, $C_{10}H_{21}CH_2C\underline{H}_2O$–)	3,212, 2,957, 2,916, 2,852, 1,738, 1,468, 1,383, 1,160, 1,075, 1,009, 722
8-1b	White solid	562.8	9.76	0.878 (t, 6H, $C\underline{H}_3(CH_2)_{13}O$–), 1.258 (m, 44H, $CH_3(C\underline{H}_2)_{11}CH_2CH_2O$–), 1.436 (d, 3H, –CH($C\underline{H}_3$)COOH), 1.625 (m, 4H, $C_{12}H_{25}C\underline{H}_2CH_2O$–), 3.820 (m, 1H, –C$\underline{H}$($CH_3$)COOH), 3.965 (m, 4H, $C_{12}H_{25}CH_2C\underline{H}_2O$–)	3,243, 2,952, 2,918, 2,850, 1,734, 1,469, 1,385, 1,151, 1,076, 1,004, 720
8-1c	White solid	619.2	9.14	0.881 (t, 6H, $C\underline{H}_3(CH_2)_{15}O$–), 1.262 (m, 52H, $CH_3(C\underline{H}_2)_{13}CH_2CH_2O$–), 1.423 (d, 3H, –CH($C\underline{H}_3$)COOH), 1.622 (m, 4H, $C_{14}H_{29}C\underline{H}_2CH_2O$–), 3.824 (m, 1H, –C$\underline{H}$($CH_3$)COOH), 3.971 (m, 4H, $C_{14}H_{29}CH_2C\underline{H}_2O$–), 4.90 (b, 1H, –N$\underline{H}$–)	3,226, 2,956, 2,916, 2,849, 1,743, 1,471, 1,389, 1,196, 1,078, 1,012, 719

and other factors have very significant effects on the stability, compressibility and phase transition behavior of the monolayer. The volume of the headgroup is the most important factor affecting the single molecule area. The increment of the length of the tail chain can improve the stability and compressibility of the monolayer. For the subphase environment, the temperature can have a significant effect on the phase transition behavior of monomolecular films; and the pH value can have an effect on single molecule area observably. For example, when the pH value is lower than the isoelectric point of the headgroup amino acid, the single molecule area is the smallest and the monomolecular film is the most dense.

8.2.2.2 Condensation of amphiphilic *N*-phosphoryl amino acids at the water–air interface

Under the primitive Earth conditions, the initial formation of early biomacromolecules may be related to the aggregation of small-molecule precursors at some ordered interfaces. Amphiphilic *N*-phosphoryl amino acids can form stable monomolecular membranes at the water–air interface. These amphiphilic molecules are well packed in the monomolecular membrane structures, and the functional headgroups are arranged in a certain orientation, which may be beneficial to the occurrence of chemical reactions. Therefore, the exploration on the reactions of amphiphilic *N*-phosphoryl amino acid in the ordered assembly system at water–air interface can supplement and improve the "co-origin" model based on *N*-phosphoryl amino acids. If this kind of model molecules is extended to the ordered interface structures, it will be very helpful and significant to understand the origin of the biological macromolecules and biofilm.

Taking di(*N*-hexadecyloxy)-*N*-phosphoryl-L-alanine (8-1c, DNHP-L-Ala) as an example, the condensation reaction of amphiphilic *N*-phosphoryl amino acid at the water–air interface was investigated. And the resulting product was analyzed by Π–*A* isotherm, and electrospray ionization-mass spectrometry (ESI-MS*n*). The related studies have shown that *N*-phosphoryl amino acids at the water–air interface can spontaneously occur the condensation reactions to form peptides in the monomolecular membranes, which has a unique dipeptide selectivity. The resulting *N*-phosphoryl dipeptide as the only product does not continue to react again to prolong the peptide chain. The possible reaction mechanism is shown in Figure 8.3.

In addition, after the condensation reaction, the supramolecular structure of the monomolecular film could be maintained. The condensation products such as *N*-phosphoryl dipeptide (DNHP-L-Ala-L-Ala) and the by-product organic phosphate (DNHP-OH) were combined in the monomolecular film through the hydrophobic interaction and the hydrogen bond of the headgroup. Based on the above experimental results, the possible mechanism of the condensation reaction of amphiphilic organic phosphate involved in quenching peptide chain growth at the water–air interface is proposed (Fig. 8.4), which could well explain the dipeptide selectivity of the interface condensation reaction.

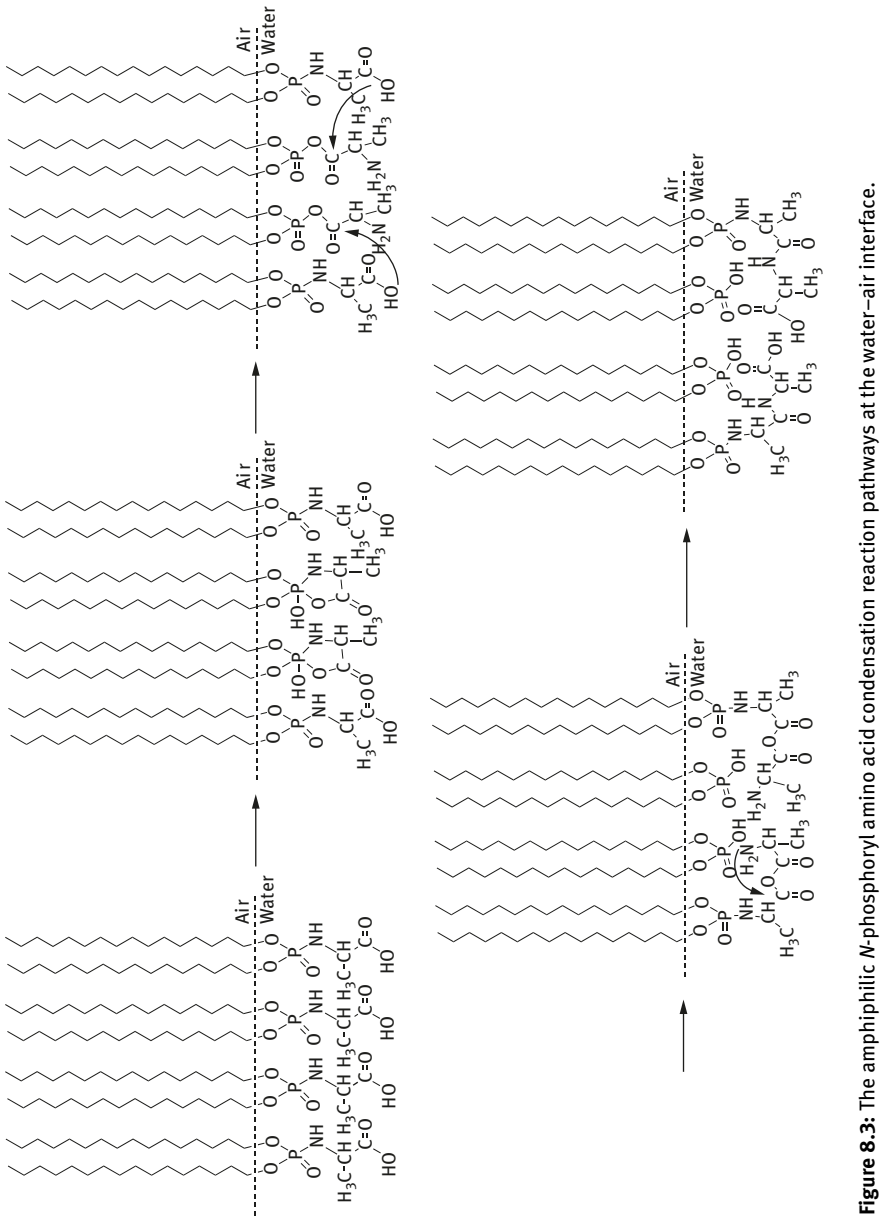


Figure 8.3: The amphiphilic *N*-phosphoryl amino acid condensation reaction pathways at the water–air interface.

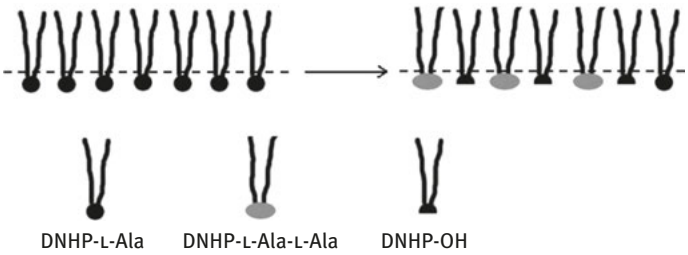


Figure 8.4: The possible mechanism of the condensation reaction of amphiphilic organic phosphate involved in quenching peptide chain growth.

8.2.3 Formation and reaction of bimolecular membranes—vesicles in water by amphiphilic *N*-phosphoryl amino acids

8.2.3.1 Formation and characterization of vesicles

Taking the compound 8-1c (DNHP-L-Ala) as an example, the self-assembly feasibility of the amphiphilic molecules into bimolecular membranes, namely vesicles, was investigated in detail. The DNHP-L-Ala prepared in our laboratory can be self-assembled into single-chamber vesicles by ultrasonic vibration and pH-controlled ultrasonic oscillating method, respectively [9, 10]. The morphology and size distribution of vesicles can be characterized by transmission electron microscopy (TEM), and the results are shown in Figure 8.5. From Figure 8.5, it can be clearly observed that DNHP-L-Ala can be assembled into a single-chamber spherical vesicle under both ultrasonic and pH-controlled ultrasonic oscillations. The size of the vesicles is about 100 nm by the TEM observation.

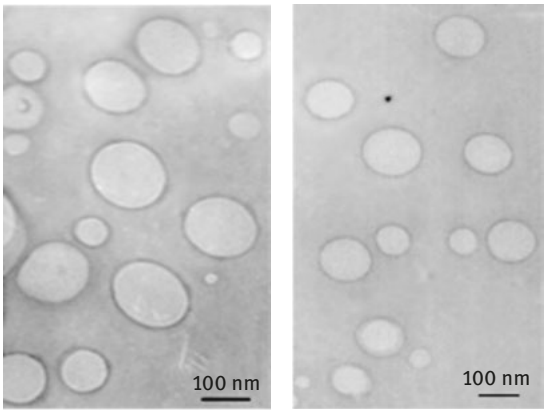


Figure 8.5: TEM images of vesicles prepared by bis(*N*-hexadecyloxy) *N*-phosphoryl alanine (compound 8-1c): (a) ultrasonic oscillation method and (b) pH control ultrasonic oscillation method.

The size and morphology of vesicles can be visually observed by TEM. In addition, the size distribution of vesicles can be detected through light scattering technology. To combine the above two technologies, the resulting vesicles can obtain more comprehensive characterization. Polydispersity is also an important property of vesicles, which is a constant in the range from 0 to 1 obtained by analyzing the particle size distribution curve. It can effectively reflect the size distribution of the vesicles. For instance, the larger the polydispersity value, the wider the particle size distribution in the system.

The vesicles prepared by the two methods with compound 8-1c were detected by laser dynamic light scattering, and the resulting particle size distribution is shown in Figure 8.6.

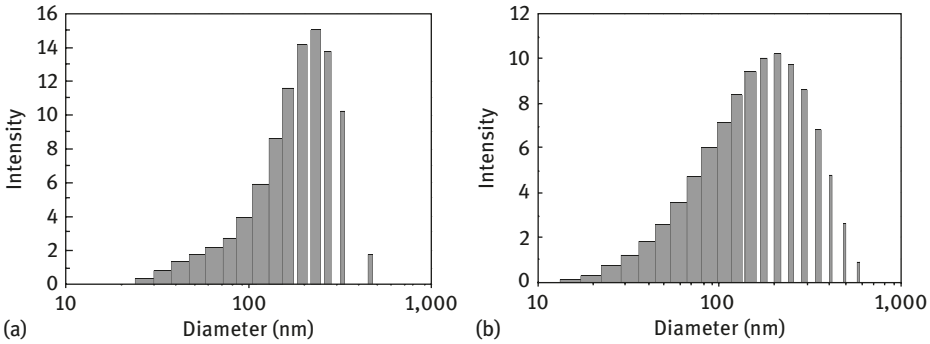


Figure 8.6: Particle size distribution of compound 8-1c (DNHP-L-Ala) vesicles: (a) ultrasonic oscillation method and (b) pH control ultrasonic oscillation method.

The average particle size and the particle size distribution of the DNHP-L-Ala vesicles obtained by the above two methods are shown in Table 8.2.

Table 8.2: Particle size distribution of compound 8-1c (DNHP-L-Ala) vesicles obtained by different preparation methods.

Methods	Average particle size (nm)	Polydispersity
Ultrasonic oscillation	212.5	0.35
pH control ultrasonic oscillation	196.0	0.38

In addition, our systematic studies indicated that the experimental dosage of *N*-phosphoryl amino acids, and the preparation and hydration of *N*-phosphoryl amino acid solid films are all key factors for the vesicle formation. The vesicles obtained by the above two methods are very similar in particle size and distribution, but the stability of the vesicles from pH-controlled ultrasonic oscillations is better than the other method. From Table 8-2, the polydispersity of DNHP-L-Ala vesicles is up to about 0.4,

indicating that the vesicles of DNHP-L-Ala have a fairly wide range of particle size distribution. Therefore, there is a great difference among the vesicles in the observed local area based on the vesicle size by the TEM and the average particle size of all vesicles obtained by the laser light scattering.

8.2.3.2 Condensation in vesicle systems

Amphiphilic molecules can align orderly in the closed bimolecular layer of vesicles. For artificial synthetic phospholipids containing specific functional groups, the vesicle bilayer can help the functional groups to have the specific orientation and tight alignment, which is conducive to the mutual reaction between them. To this end, the study on the reaction of amphiphilic *N*-phosphoryl amino acid in the vesicle bimolecular layer was carried out in our laboratory.

The vesicles of DNHP-L-Ala were prepared by pH-controlled ultrasonic oscillating method with a target pH of 8.0. Compound 8-1c vesicle dispersion liquid was oscillated and incubated at 40 °C for 48 h. After that, the resulting vesicle dispersion liquid was freeze-dried, and then the lyophilized solid product was dissolved in methanol and analyzed by ESI-MS and ESI-MS/MS.

First, the vesicle dispersion liquid after incubation was analyzed by ESI-MS in negative mode. From the mass spectrum shown in Figure 8.7, it was found that the molecular ion peak of compound 8-1c has completely disappeared, whereas the two new ion peaks have emerged. One is the molecular ion peak of the phosphoryl dipeptide product (*N*-phosphoryl-L-Ala-L-Ala) at *m*/*z* 687.7, and another one corresponds to the hydrolysis product DNHP-OH at *m*/*z* 545.8.

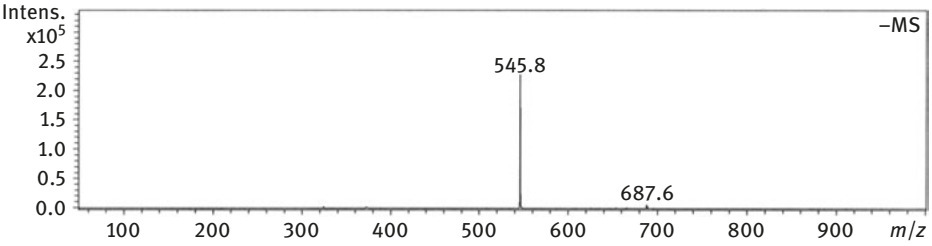


Figure 8.7: ESI-MS spectrum of compound 8-1c after vesicle incubation (in negative ion mode).

In order to confirm its structure further, the molecular ion peak of the dipeptide product at *m*/*z* 687.6 was analyzed by ESI-MS/MS (Figure 8.8). Besides, the standard DNHP-L-Ala-L-Ala was also dissected by ESI-MS/MS. Through the comparison of Figure 8.8 and the fragment pattern coming from the ESI-MS/MS spectrum of the standard, it was found that the dipeptide product and the standard have the identical fragment pattern. That means the interface condensation product is DNHP-L-Ala-L-Ala exactly.

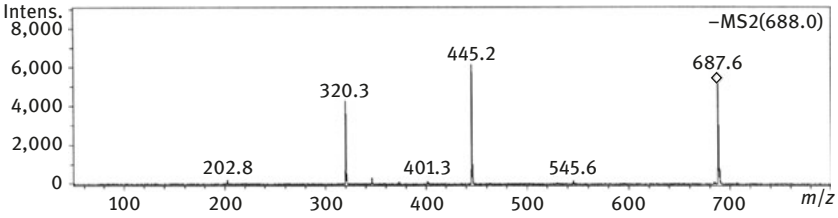


Figure 8.8: ESI-MS/MS spectrum of dipeptide products DNHP-L-Ala-L-Ala.

In addition, similar to the condensation reaction at the water–air interface, the condensation reaction of DNHP-L-Ala in vesicles can also produce the phosphorylated dipeptides without any longer polypeptide than dipeptide. That means the condensation reaction within the vesicles also has dipeptide selectivity.

The situation in the vesicle bimolecular layer is similar to that in the water–air interface. Phosphoryl dipeptide and organic phosphate byproducts produced by the condensation reaction are also amphiphilic. After the reaction, the hydrophobic tail chains of the phosphoryl dipeptide product and the organic phosphate byproduct are "locked" in the bilayer under the influence of the hydrophobic interaction, maintaining the closed bilayer structure of the vesicles. The reaction byproduct amphiphilic phosphates are involved in the construction of a closed bilayer membrane structure, which intervenes between phosphodiester products, spatially interfering with the continued attack of the phosphodiester dipeptides, thereby halting the continued growth of phosphoryl dipeptides and making the reaction to have unique dipeptide selectivity. The mechanism of the byproduct amphiphilic phosphate involved in terminating peptide chain growth in the vesicle system can be illustrated graphically using the cartoon diagram in Figures 8.9.

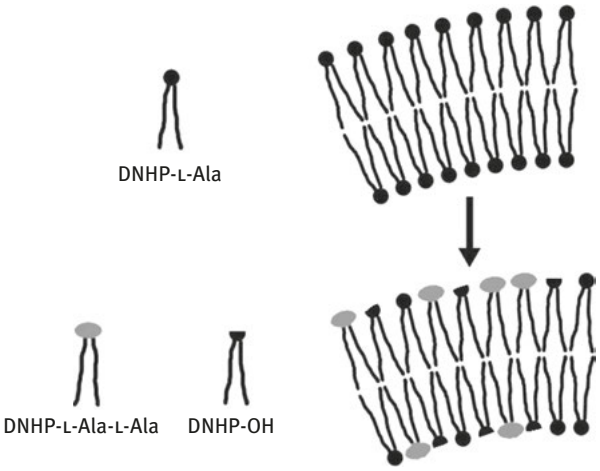


Figure. 8.9: The mechanism of amphiphilic organic phosphate involved in quenching the peptide chain growth in the vesicle system.

The above-mentioned results indicate that amphiphilic *N*-phosphoryl amino acids can not only be assembled into monomolecular films at the water–air interface, but can also effectively self-assemble into bimolecular membrane – vesicles – in the aqueous phase and can be selectively condensed into dipeptides in both the above two media. These research results can further complement and improve the "co-origin" model based on *N*-phosphoryl amino acids, which will provide a novel insight on the origin of biomacromolecules and biofilm in ordered interfaces, and are of great scientific significance for understanding the molecular mechanism of chemical origin of primitive cell membrane.

8.3 Evolution of the protocell membrane

Cell generation and evolution have been going through billions of years. Scientists have synthesized a variety of phospholipid membrane molecules to study the property difference between the artificial membrane and the cell membrane. As the research moves along, the evolutionary theory of cell membrane has been improved gradually. In 2004, Professor J. W. Szostak, reported that the cell membrane also has the ability to trigger its evolution [11]. The primitive cell membrane could be a simple linear phospholipid molecule without the integrated properties of modern cell membranes, such as regulating osmotic pressure. However, in a very dilute phospholipid solution, the simple structure unit of the linear molecular membrane will be gradually replaced by free phospholipids, so as to achieve the enrichment of phospholipid and cell membrane evolution (Figure 8.10). Besides, based on the above research report, it implies that the amphiphilic phosphorylated amino acid molecules are likely to play an important role on the evolution of cell membranes.

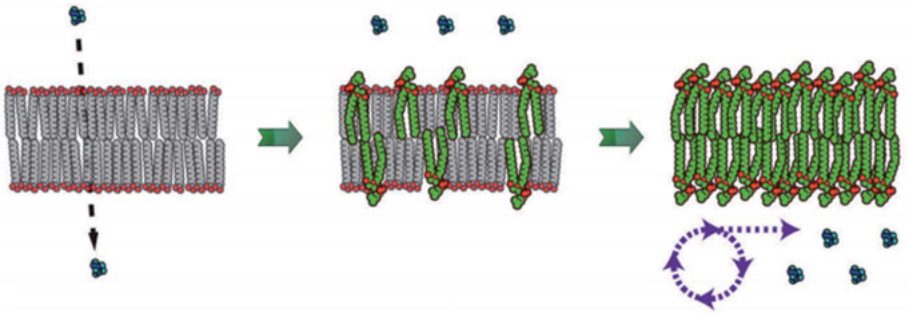


Figure 8.10: The mechanism of the evolution of the cell membrane (gray represents a linear molecule, and green represents phospholipids) [11]. Reprinted with permission [PNAS].

8.4 *N*-Phosphoryl amino acids and the origin of life

8.4.1 Synergistic effect of membrane, nucleic acid and protein

Almost all of the life is composed of cells. Cell membrane has the great significance for life. Cell membrane provides a relatively stable living environment, and the integrated functions produced by the gradual evolution of it are also more conducive to the survival and activities of life. Biologists also found that cell membrane and all kinds of life activities within membrane have a synergistic effect [12] between each other. The membrane dependence of life and the promotion of membrane environment for life activities conjointly contribute to the emergence of primitive cells.

8.4.2 The coevolution theory of nucleic acid, protein and cell membrane

For the origin of life, "RNA world" theory emphasizes that RNA, which exhibits the most basic properties required for life, such as self-replication, proliferated before the evolution of DNA. At the same time, the "RNA world" theory also believes that "the protein structure is so complicated, that the primitive life does not have the ability to synthesize the protein molecules. Therefore, the evolution of primitive life impossibly has the participation of protein." However, in 2001, Professor J. W. Szostak boldly proposed the idea of "synthetic life," and pointed out that the most primitive life requires not only genetic material but also the original cell membrane and functional molecules that regulate membrane growth [13].

With the deep studies on the chemical evolution of life, some small molecules with various biological activities, such as small peptides, are gradually discovered. Among them, the most prominent and widely regarded functional peptide is seryl–histidine dipeptides (Ser–His). In 2009, Professor P. L. Luisi found that Ser–His could catalyze the formation of peptide bonds (Figure 8.11), such as the formation of the phenylalanyl–leucine dipeptides [14]. In 2013, Professor Szostak's research team found that, in the vesicle system, Ser–His can regulate the growth of the cell membrane through in situ catalyzing the formation of the phenylalanyl–leucine dipeptides [15] (Figure 8.12). The above experimental results indicate that small peptides really play a very important role on the chemical evolution process of life, further providing strong evidences for supporting the coevolution theory of nucleic acid, protein and cell membrane.

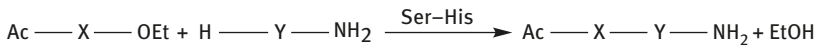


Figure 8.11: Catalytic reaction of seryl–histidine dipeptide for the formation of peptide bonds.

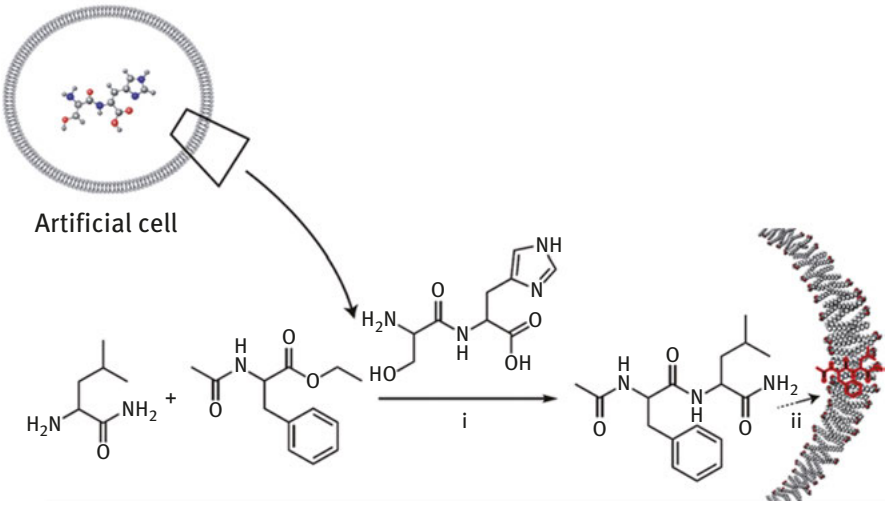


Figure 8.12: A schematic representation of seryl–histidine dipeptide catalyzing peptide bond formation, in which the produced dipeptide promotes the proliferation of artificial cell membranes [15]. Reprinted with permission [Springer Nature].

In Figure 8.11, X and Y represent two different amino acid residues in the reaction formula. For example, X is the phenylalanine residue and Y is the leucine residue. Then phenylalanyl–leucine dipeptide is obtained by the catalysis of Ser–His.

In 2015, Professor Sutherland achieved the one-pot synthesis of the uridine-2′,3′-cyclic phosphate, amino acids and liposomes precursor by hydrogen cyanide (HCN) and its related derivatives, H_2S as a reducing agent under the ultraviolet radiation, with the cyclic catalysis of Cu(I)–Cu(II) [16]. After publishing the research work, it has received extensive attention from scientists, who published papers in *Science* in that month, as well as received high praise about this "three in one" hypothesis of the origin of life [17].

However, the above primitive cells hypothesis of nucleic acid, protein and cell membrane all ignore the participation and regulation role of "phosphorus" in the chemical evolution process of life. *N*-Phosphoryl amino acids can be used as a model for "three in one" primitive cells evolution since it can associate the emergence of oligonucleotides, oligopeptides and phospholipids with each other.

References

[1] Davis B D. On the importance of being ionized [J]. *Arch. Biochem. Biophys.*, 1958, 78: 497–509.
[2] Deamer D W. The first living systems: a bioenergetic perspective [J]. *Microb. Mol. Biol. Rev.*, 1997, 61: 239–261.
[3] Segre D, Ben-Eli D, Deamer D W et al. The lipid world [J]. *Origins Life Evol. Biosph.*, 2001, 31: 119–145.

[4] Blocher M, Liu D J, Walde P et al. Liposome-assisted selective condensation of α-amino acids and peptides [J]. *Macromolecules*, 1999, 32: 7332–7334.
[5] Ferris J P, Hill A R, Liu R H. Synthesis of long prebiotic oligomers on mineral surfaces [J]. *Nature*, 1996, 381: 59–61.
[6] Son H L, Suwannachot Y, Bujdak J et al. Salt-induced peptide formation from amino acids in the presence of clays and related catalysts [J]. *J. Inorg. Biochem.*, 1996, 61: 69–78.
[7] Wang H Y, Li Y M, Xiao Y. Condensation properties of vesicles formed from an amphiphilic N-phosphorylamino acid [J]. *J. Colloid Interf. Sci.*, 2005, 287: 307–311.
[8] Li Y M, Zhao Y F. Effects of N-phosphoproline on the phospholipid of human erythrocytes membrane [J]. *Spectro. Lett.* 2000, 33: 653–659.
[9] Engberts J B F N, Hoekstra D. Vesicle-forming synthetic amphiphiles [J]. *BBA-Rev. Biomembranes*, 1995, 1241: 323–340.
[10] Fuhrhop J H, Mathieu J. Routes to functional vesicle membranes without proteins [J]. *Angew. Chem. Int. Ed. Engl.*, 1984, 23: 100–113.
[11] Budin, I. Szostak, J. W. Physical effects underlying the transition from primitive to modern cell membranes [J]. *Proc. Natl. Acad. Sci.* U S A, 2011, 108 (13): 5249–5254.
[12] Chen I A, Salehi-Ashtiani K, Szostak J W. RNA Catalysis in model protocell vesicles [J]. *J. Am. Chem. Soc.*, 2005, 127: 13213–13219.
[13] Szostak J W, Bartel, D P, Luisi P L. Synthesizing life[J]. *Nature*, 2001, 409: 387–390.
[14] Gorlero M, Wieczorek R, Adamala K et al. Ser-His catalyses the formation of peptides and PNAs [J]. *Febs. Lett.*, 2009, 583(1): 153–156.
[15] Adamala K, Szostak J W. Competition between model protocells driven by an encapsulated catalyst. [J]. *Nat. Chem.*, 2013, 5: 495–501.
[16] Patel B H, Percivalle C, Sutherland J D et al. Common origins of RNA, protein and lipidprecursors in a cyanosulfidicprotometabolism [J]. *Nat. Chem.*, 2015, 7: 301–307.
[17] Service R F. Origin of life puzzle cracked [J]. *Science*, 2015, 347: 6228.

9 The potential evolution prototype of modern enzyme: Discovery of seryl–histidine dipeptide and its function

Introduction: Seryl–histidine dipeptide (Ser–His) is a kind of magical dipeptide, which has been regarded as the shortest peptide with reversible catalytic activity. It not only hydrolyzes protein and DNA, but also catalyzes the formation of peptide bond and phosphodiester bond. In addition, its hydrolysis activity has a general applicability for substrate. The aforementioned features are similar to that of the modern enzyme. Therefore, it is reasonable to believe that Ser–His, as the miniprotease and nuclease, can be regarded as the potential candidate for prototype of modern hydrolytic enzyme.

9.1 Mini-activating enzyme: Ser–His

Proteases are any enzyme that can hydrolyze peptide bonds into amino acids. Almost one-third of proteases can be categorized as serine proteases, which was intrinsically distinguished by the presence of the Ser–His–Asp catalytic triad. In the three-dimensional structure of serine proteases, amino acid residues – serine (Ser) and histidine (His) – are not adjacent in the primary sequence, but they are in sufficiently close proximity to function together in the active sites [1–6]. In the hydrolysis process, the hydroxyl group of serine side chain usually takes as the nucleophilic reagent to participate the reaction. Histidine is the enhancer for hydrolysis process, because the imidazolyl group of its side chain can be either a donor or a receptor of protons so as to take part in the reaction with serine cooperatively [7–9]. Although Asp–His–Ser catalytic triad in the active center of serine protease is very important, Ser–His catalytic diad also exhibits effective hydrolysis activity [6, 10].

About 20 years ago, when we investigated the biological activity of *N*-phosphoryl serine, Ser–His, a especial dipeptide with different hydrolysis activities, was discovered, whose molecular structure is shown in Figure 9.1. Till now, Ser–His is the reported smallest functional peptide with various hydrolysis activities. It not only hydrolyzes DNA, but also cleaves protein and carboxylic ester. In addition, Ser–His can also catalyze the formation of peptide and phosphodiester bonds, which resemble the modern enzyme with the microscopic reversibility. Consequently, Ser–His, as a minienzyme, could be the prototype of modern hydrolase during the chemical evolution process.

https://doi.org/10.1515/9783110562552-009

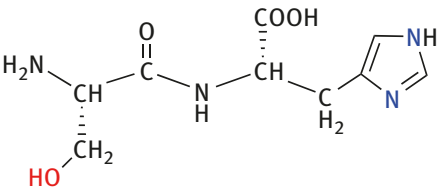


Figure 9.1: The molecular structure of Ser–His dipeptide.

9.2 The cleavage activities of Ser–His on DNA

9.2.1 Discovery of the hydrolysis activity of Ser–His on DNA

When we studied the interaction of *N*-phosphoryl-serine with DNA, it was found that the old *N*-phosphoryl-serine solution in a saturated histidine buffer displayed the cleavage activity on DNA, but not the fresh one. Subsequently, it was clarified that Ser–His takes charge of the cleavage activity. However, why does Ser–His exist in the old saturated histidine solution? The answer is *N*-phosphoryl-serine could react with histidine to produce *N*-phosphoryl-serine–histidine, which will release Ser–His in the old solution by the hydrolysis of labile P–N bond. The related mechanism is illustrated in Figure 9.2. Ser–His can cleave DNA in a wide pH range from 5 to 9. The optimal pH value is 6 at an incubation temperature 37 °C, which is approximate to the pK_a of imidazolyl of histidine side chain (pK_a = 6). For example, both linear bacteriophage λ-DNA and circular plasmid DNA pBR322 were cleaved into smaller fragments of heterogeneous sizes, after 72 h incubation with Ser–His. Besides, the cleavage rate depends on the incubation temperature. The higher temperature such as 50 °C has a faster cleavage rate than that at 37 °C. The cleavage selectivity of Ser–His on DNA could not be observed by ^{32}P isotope autoradiography [11].

To eliminate the possibility of DNA cleavage due to nucleases or contamination of metal ions, Ser–His samples used in the related experiments were all filter-sterilized or autoclaved. The control experiments that DNA incubated with Ser–His in the presence or absence of EDTA were performed and analyzed by agarose gel. The results indicated that irrespective of whether EDTA was present or not, the treated Ser–His showed obvious cleavage activity for DNA. When Cu^{2+} or Fe^{2+} was added to the DNA solution without Ser–His, the DNA cleavage activity could not be observed irrespective of whether EDTA is present or not. Besides, Ser–His obtained from various sources and the one purified by high-performance liquid chromatography all exhibited cleavage activity for DNA. This means that chemical impurities in Ser–His samples did not give rise to the cleavage activity.

The aforementioned results were summarized in Table 9.1, which clearly shows that Ser–His has the distinct cleavage activity for DNA.

It should be noted that the cleavage activity of Ser–His on DNA is much lower than DNase. Specifically, the cleavage activity of 1.21 μg/μL Ser–His is equal to 1/1,000 of 7×10^{-3} unit/μL RQ1 DNase at 37 °C.

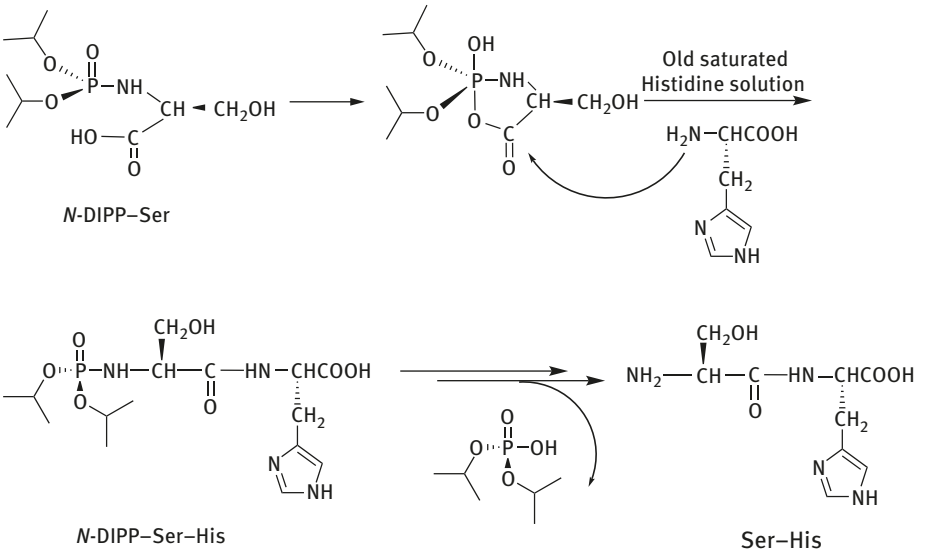


Figure 9.2: The peptide formation mechanism of *N*-phosphoryl-serine (*N*-diisopropyl phosphoryl Ser, *N*-DIPP–Ser) in old saturated histidine solution.

Table 9.1: The relevant experimental results about cleavage activity of Ser–His on DNA [11].

Reaction condition[a]	DNA cleavage activity[d]
Britton–Robinson (B–R) buffer	–
Ser–His[b] (filtered)+B–R	+
Ser–His[b] (autoclaved)+B–R	+
Ser–His[b] ± 1 mM EDTA +B–R	+
Ser–His[b] +B–R at 65 °C	+
$FeSO_4$[c] ± EDTA +B–R	–
$CuSO_4$[c] ± EDTA +B–R	–

[a] All reactions were performed at 37 °C (or 65 °C) in 40 mM Britton–Robinson buffer (B–R buffer, pH 6.0) with 20 ng/mL λ-DNA in a total volume of 20 mL.
[b] Ser–His purchased from three different sources was used in this study.
[c] The $FeSO_4$ and $CuSO_4$ solutions were each 1 and 10 mM.
[d] "+" stands for with the cleavage activity, and "–" stands for without the cleavage activity.

9.2.2 The cleavage mechanism of Ser–His on DNA and the pivotal role of functional groups in Ser–His

T4 DNA ligase can ligate free 3′-hydroxyl groups and 5′-phosphates of oligonucleotides to form the longer oligonucleotides. Therefore, the cutting fragments generated from the cleavage experiment of λ-DNA with Ser–His were then incubated

with T4 DNA ligase. It was found by agarose gel [11, 12] that cleavage products could be ligated into the longer oligonucleotides. This indicated that these terminal groups of cleavage products are free 3′-hydroxyl groups and free 5′-phosphates, which are consistent with hydrolysis of phosphodiesters. Hence, the cleavage mechanism of Ser–His is the hydrolysis mechanism, but not a free-radical mechanism.

To evaluate the cleavage activity of related oligopeptides and the role of the functional groups of Ser–His in DNA cleavage, different amino acid residues were used instead of Ser or His in Ser–His dipeptide or added internally or to the N- or C-terminals of Ser–His. The cleavage activities disappeared when Ser was replaced by any other amino acid except cysteine (Cys) or threonine (Thr). Besides, if the histidine residue of Ser–His was replaced by lysine (Lys) or arginine (Arg) with positive charge in side chains, the cleavage activities were also lost. However, if alanine (Ala) replaces the histidine residues, the cleavage activities survive, but were weaker than those of Ser–His [13]. His–Ser, a dipeptide with the same chemical composition of Ser-His but in a reverse sequence, is also inactive. The cleavage activities of Ser–His are reduced or lost when an amino acid is added to its N-terminus but are retained when one or more amino acids are added to its C-terminus. It was also found that amino acids could be added in the middle of Ser and His residues without abolishing the cleavage activities.

The aforementioned experimental results indicated that the hydroxyl group of N-terminal amino acid residues side chain (Ser, Thr) or sulfhydryl in Cys is an essential functional group for the cleavage activities on DNA. The imidazole group of histidine is just a kind of enhancer to promote the cleavage function of the hydroxyl group of Ser. Up to now, there have been so many reports that show that the imidazole groups have an important catalytic or provocative role in the natural enzymatic reaction [14–16]. This means that the imidazole group is very helpful in promoting the hydrolysis function of natural enzyme, though it is not the essential group. This fact is consistent with the results of our research.

In addition, a series of compounds with similar structures of Ser-His, such as Boc–Ser–His, Ser–His–OMe, Ser–OMe and Thr–OMe, were treated to investigate the cleavage activity on DNA [13, 17]. Boc–Ser–His is inactive; however, Ser–His–OMe, Ser–OMe and Thr–OMe showed some cleavage activities, but weaker than that of Ser–His. It indicated that the C-terminal of Ser–His is not important for DNA cleavage. The esterification of C-terminal decreases the dissolvability of Ser–His in aqueous phase, which may result in the reduction of activities. It should be noted that all four different hydramines, namely, ethanolamine, diethanolamine, triethanolamine and isopropanolamine, exhibited some cleavage activities in our previous research. These results indicated that both the amino group of N-terminal and hydroxyl group are the key functional groups for cleavage activities. Furthermore, they should be close to each other.

All the results are summarized in Figure 9.3.

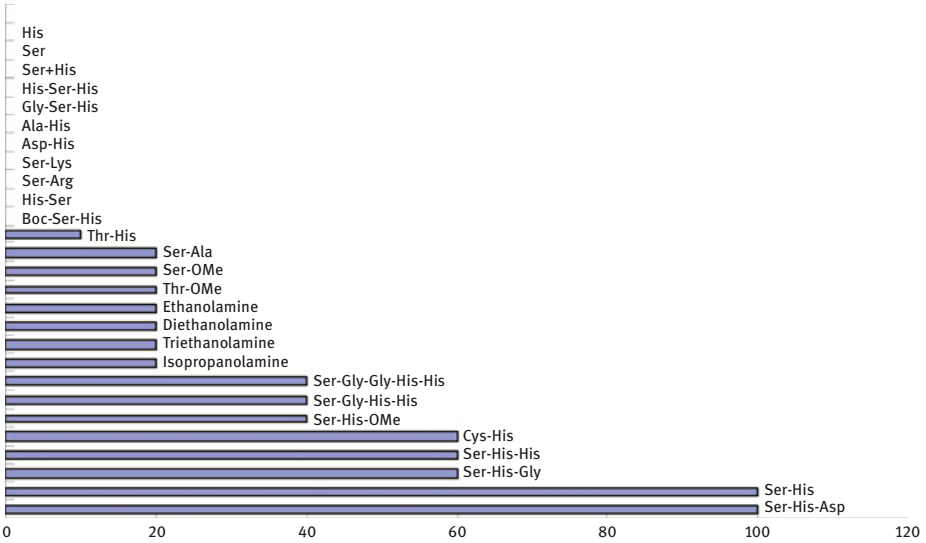


Figure 9.3: The cleavage activity of Ser–His, the related oligopeptides and compounds with similar structures [11, 12, 17].

On the basis of the evaluation of the role of the functional groups of Ser–His, Ser–Hismine amide (a similar structure of Ser–His without carboxyl), was synthesized and the DNA cleavage activities were also observed [18, 19].

9.2.3 Molecular modeling for investigating DNA cleavage activity of Ser–His

A series of three-dimensional models of dipeptides (such as Ser–His) and 5'-TpTpdC-3' complexes have been constructed using the molecular docking program FlexiDock encoded in Sybyl [20–22]. From the optimal conformation of these complexes, it was found that the distance between the oxygen atom of the hydroxyl group on Ser–His and the oxygen atom of phosphate group has the key effect on the cleavage activities. For Ser–His, the mentioned distance is 3.7 Å, which is within a contactable range for oxygen atom conducting nucleophilic attack to the phosphoric ester. It is beneficial to cleave the ester bond through forming a pentacoordinate phosphotriester transition state, and this subsequently results in the DNA cleavage.

Another key structure feature in the interaction complexes is that Ser–His can simultaneously bind to two neighboring phosphates in the DNA backbone. For Ser–His–5'-TpTpdC-3', the Ser residue binds to a phosphate through H-bonds and electrostatic interaction, while the His binds to the adjacent phosphate (Figure 9.4). These H-bonds make Ser–His and 5'-TpTpdC-3' together to realize the nucleophilic attack of the hydroxyl group of Ser residue to the phosphate.

So far, combining all the experimental evidences and molecular modeling results, the DNA cleavage mechanism of Ser–His is postulated as shown in Figure 9.4.

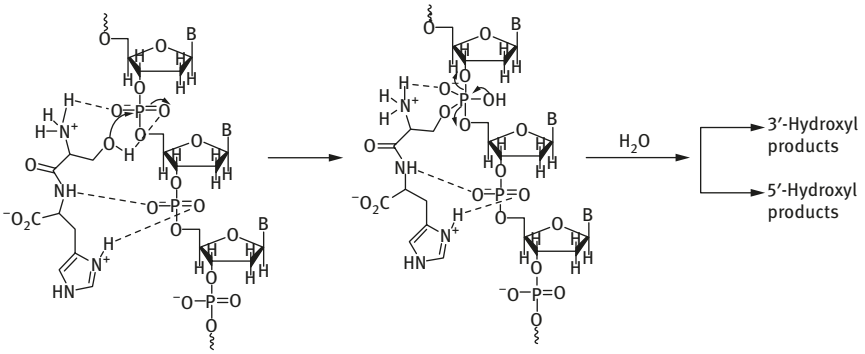


Figure 9.4: Proposed mechanism for DNA cleavage by Ser–His. Dash lines represent H-bonds between Ser–His and oligonucleotide [17]. Reprinted with permission [Springer Nature].

9.3 The cleavage activities of Ser–His on proteins

As two most important kinds of biomacromolecule, protein and nucleic acid that take charge of the generality and specialty of each living system are the key research objects. As mentioned earlier, serine and histidine are the pivotal amino acid residues in the catalytic center of serine protease. Therefore, the studies on the cleavage activities of Ser–His on proteins also have important theoretical significance and potential application value.

9.3.1 The discovery of the cleavage activities of Ser–His on proteins

Under the similar conditions as DNA cleavage experiments, Ser–His can also cleave proteins by hydrolysis mechanism, such as bovine serum albumin (BSA) [11, 23], green fluorescent protein (GFP) [24], and cyclophilin A (CyPA) [25]. Similarly, with DNA cleavage, proteins could be easily cleaved by Ser–His under the incubation condition, with an optimal pH = 6. About the cleavage temperature, the higher the temperature is, the faster the cleavage rate is. However, the temperature could not exceed the denatured temperature of protein. In particular, 50 °C is better for BSA and GFP, but not for CyPA. The temperature 50 °C will result in the denaturation of CyPA, where optimal cleavage temperature for CyPA is 37 °C.

Whether the protease contamination gives rise to the protein cleavage in the incubation experiments of Ser–His? To answer this question, before all the related experiments, Ser–His was filter-sterilized or autoclaved to inactivate protease contamination.

In addition, for all the cleavage experiments, two parallel tests were simultaneously carried out. One is treated with Ser–His that acts as the experimental group and the other is without Ser–His that acts as the control group. Results indicated that only the experimental group with Ser–His exhibited distinct cleavage activities, but not the control group. When the protease inhibitor phenyl methyl sulfonyl fluoride was added to the cleavage experimental group, the same result could be observed. This means that Ser–His definitely has the protein cleavage activity. In addition, it should be noted that the protein cleavage activity of Ser–His is weaker than that of modern protease, such as protease K. In B–R buffer with pH = 6.5–7.5, the incubation temperature is 50 °C, incubation time is 24 h and the cleavage activity of 1 mmol Ser–His was close to 0.33×10^{-4} mmol protease K activity [24].

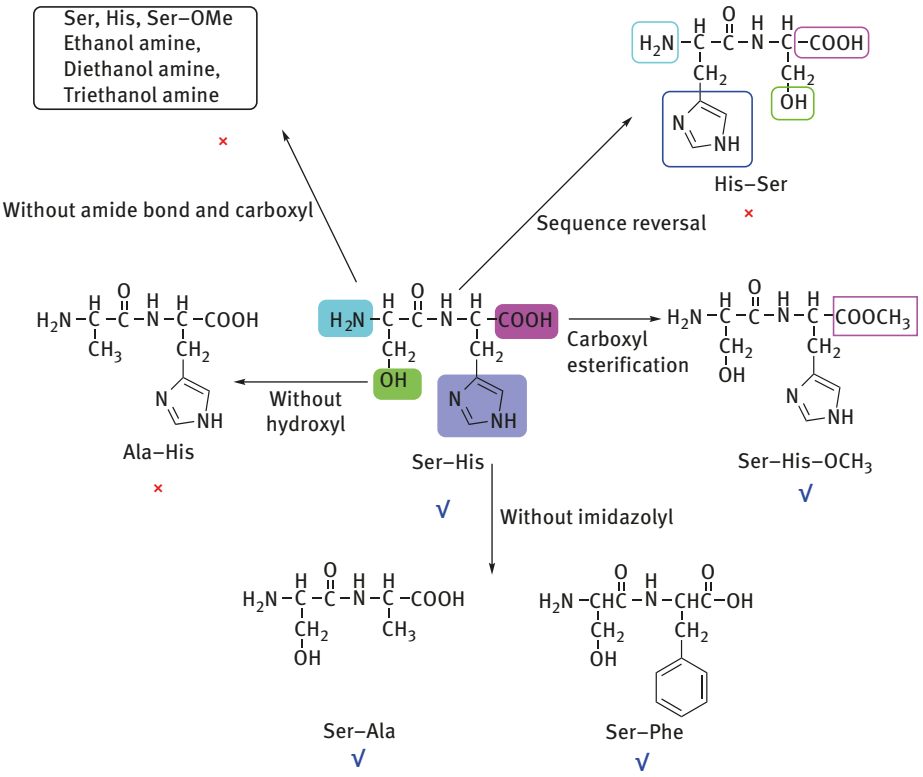
9.3.2 The effect of different buffers on the cleavage activities

To investigate the effect of different buffers on the cleavage activities, five buffer conditions (pH value = 6) were used in the BSA cleavage experiments of Ser–His under 60 °C incubation for 24 h. The buffer systems used here were B–R buffer, citric acid–citric acid trisodium buffer, disodium hydrogen phosphate–sodium dihydrogen phosphate buffer, Bis–Tris buffer and the system without any buffer additive (the Ser–His solution itself can maintain a pH value about 6), respectively. The related results indicated that among the common biochemical buffers, B–R and phosphate buffer could enhance the cleavage activity, while citric acid–citric acid trisodium buffer and Bis–Tris inhibited the cleavage reaction. Without any buffer additive, BSA after incubation with Ser–His could be cleaved into some fragments by SDS-page observation, and the original BSA bond faded a little. However, when compared with the B–R or phosphate buffer system, this cleavage activity was unapparent [26].

9.3.3 The role of functional groups of Ser–His in the cleavage activities

To understand the role of the functional groups of Ser–His during the cleavage reaction, a series of dipeptides and amino acids with the similar Ser–His structures were incubated with BSA to explore the differences of cleavage activities [26]. The dipeptides and amino acids tested here were His–Ser, Ser–Ala, Ala–His, serine, histidine and Ser–OMe. Besides, the protein cleavage activities of ethanol amine, diethanol amine and triethanol amine were all tested in detail. The relevant experimental results showed that the following: (1) Ala–His is inactive for BSA, but Ser–Ala is active with weaker effect. It means that the hydroxyl group of Ser–His is the essential functional group for the cleavage activity. The imidazole group is the enhancer to promote the cutting reaction. The aforementioned results are consistent with DNA cleavage activities of Ser–His. (2) Serine, Ser–OMe, ethanol amine, diethanol amine and triethanol amine are

inactive for BSA. The results are not an exact match with DNA cleavage experiments. In view of the weak cleavage activity of Ser–Ala, it could be concluded that the amide bond and the carboxyl group have some contribution for BSA cleavage. These two functional groups may be beneficial to make Ser–His and substrate together to help the hydroxyl group of Ser–His to boost the next nucleophilic attack. (3) Separate both Ser and His are inactive for BSA, which further indicated that Ser–His is responsible for the BSA cleavage, not Ser or His that could be released from the hydrolysis of Ser–His. (4) His–Ser with the reverse sequence of Ser–His is inactive for BSA, which is also identified in DNA cleavage experiment. All the research results are shown in Figure 9.5.



(√ stands for active, and × stands for inactive)

Figure 9.5: The BSA cleavage activities of Ser–His and related compounds.

9.3.4 The study of the interaction of Ser–His with the substrate proteins

As the initial step of the protein cleavage process, the noncovalent interaction between Ser–His and proteins is crucial for the protein cleavage activity. The study on the noncovalent interaction between Ser–His and substrate proteins will be helpful for

further understanding of the cleavage mechanisms. In our previous research, we took CyPA as a substrate protein to evaluate the noncovalent interaction with Ser–His using a combination of nuclear magnetic resonance (NMR) spectroscopy and molecular modeling approach [25]. Two independent Ser–His binding sites on CyPA (Figure 9.6) were detected using ^{15}N–^{1}H heteronuclear single-quantum coherence (HSQC) spectra. Each binding site binds one Ser–His molecule. Dissociation constants, K_{d_1} and K_{d_2}, were estimated to be 2.07 and 6.66 mmol/L, respectively, indicative of the weak noncovalent interaction between Ser–His and CyPA. The results implied that the cleavage activity of Ser–His on CyPA is weak, which was proved by SDS-page results shown in Figure 9.7.

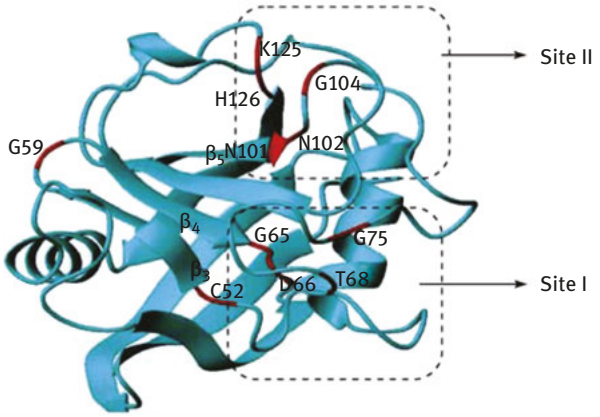


Figure 9.6: Ribbon diagram of the three-dimensional structure of CyPA with two noncovalent interaction sites of Ser–His in solution. Residues with significant chemical shift changes (red) are labeled with single letter code. Reprinted with permission [Springer Nature].

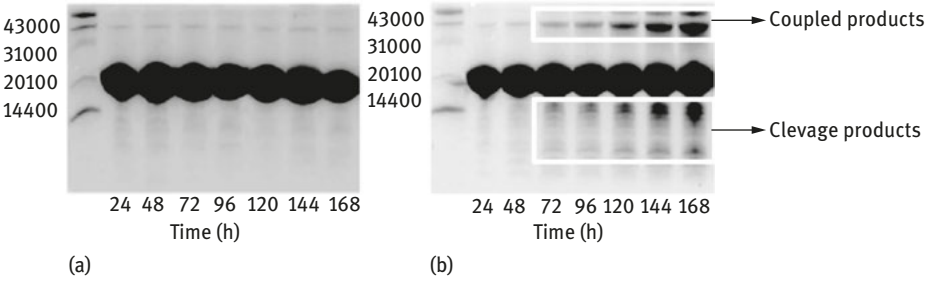


Figure 9.7: The cleavage activity experiment of Ser–His on CyPA[25]. Reprinted with permission [Springer Nature]: (a) CyPA (0.5 mmol/L) without Ser–His and (b) CyPA (0.5 mmol/L) cleaved by Ser–His (70 mmol/).

On the basis of molecular modeling results, it can be concluded that both the α-amino and hydroxyl groups of Ser–His are crucial function groups for the noncovalent interaction between Ser–His and CyPA or for the further activity of protein cleavage. The imidazole group of Ser–His (Figure 9.8b) or the amide bond from the substrate

(Figure 9.8a) can be an assistant group to increase the nucleophilicity of hydroxyl group of Ser–His. The aforementioned results are consistent with the research about the role of functional group of Ser–His in the cleavage process.

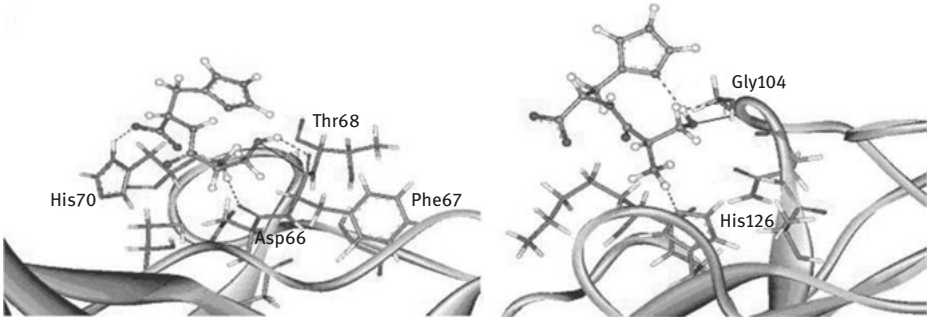


Figure 9.8: The docked structure of Ser–His binding to CyPA via either site I (a) or site II (b) on CyPA. Ser–His is displayed with balls and sticks, and the amino acid residues in the "binding pocket" of CyPA are displayed with sticks. H-bonds are marked by dash lines. Reprinted with permission [Springer Nature].

9.3.5 The cleavage activities of Ser–His on carboxylic ester

Serine protease not only hydrolyzes peptide bonds of proteins, but also cleaves the carboxylic ester. Like serine protease, Ser–His also hydrolyzes carboxylic ester, such as *p*-nitrophenyl acetate (*p*-NPA). At room temperature, *p*-NPA was incubated with Ser–His in the B–R buffer. Rates of *p*-NPA cleavage were measured by monitoring changes in the optical density at 400 nm (OD_{400}) of each sample over the time. When the incubation time was prolonged, OD_{400} value of *p*-NPA was growing linearly. Besides, the changing rate of *p*-NPA OD_{400} value depends on the concentration of Ser–His, incubation temperature and buffer pH value. These experimental phenomena indicated that *p*-NPA could be gradually hydrolyzed into *p*-nitrophenol by Ser–His [11].

9.3.6 The functional reversibility of Ser–His on the substrate

When we investigated the cleavage activity of Ser–His on CyPA, it was found that besides cleavage fragments, some bonds with larger molecular weight than that of CyPA were also observed by SDS-page experiments (Figure 9.7). In addition, the longer the incubation time, the more obvious are these bonds on SDS-page. We speculate these bonds with the larger molecular weight than that of CyPA correspond to the coupled products.

We know that protease can catalyze the formation of peptide bonds [27]. How about Ser–His? On the basis of the microscopic reversibility principle, we speculate that Ser–His

also has the function reversibility, which means that Ser–His not only hydrolyzes peptide bonds, but also catalyzes the formation of peptide bonds. In 2009, Professor Pier Luigi Luisi reported that Ser–His could catalyze the formation of peptide bonds and peptide nucleic acid (PNA) [28]. The discovery is an indirect proof that proves that the couple products could be obtained during the Ser–His incubation experiment with CyPA.

In 2013, Professor Jack W. Szostak at Harvard Medical School (the Nobel Prize winner in 2009 for physiology or medicine) simulated the protocell growth with the Ser–His as a catalyst for the formation of peptide bonds. He also discovered that the catalytic process could be promoted in closed fatty acids vesicles [29]. These works adequately indicate that Ser–His has a function reversibility for substrate proteins.

9.4 Ser–His: the evolution prototype of modern enzyme

Ser–His is the smallest reported peptide with multiple biological activities. It can hydrolyze DNA, protein and *p*-NPA within a wide pH range. The detailed biological activities of Ser–His and the related compounds are summarized in Table 9.2 [11, 14, 17, 19, 26].

Table 9.2: Multiple biological activities of Ser–His and the related compounds.

Entry	Ser–His and analogue	Substrate			Entry	Ser–His and analogue	Substrate		
		DNA	Protein	*p*-NPA			DNA	Protein	*p*-NPA
1	Ethanol amine	+	–	—	**17**	Cys–His	+	+	—
2	Diethanol amine	+	–	—	**18**	Thr–His	+	+	—
3	Trienthanol amine	+	–	—	**19**	Asp–His	–	–	—
4	Isopropanol amine	+	—	—	**20**	Ala–His	–	–	—
5	Ethamine +ethanol	–	—	—	**21**	Ser–His–Asp	+	+	—
6	Ser	–	–	—	**22**	Ser–Gly–His–His	+	+	—
7	His	–	–	—	**23**	Ser–Gly–Gly–His–His	+	+	—
8	Ser + His	–	–	—	**24**	His–Ser	–	–	—
9	Ser–His	+	+	+	**25**	Ser–His + Fe^{2+}	+	+	—
10	Ser–Ala	+	+	—	**26**	Ser–His + Cu^{2+}	+	–	—
11	Ser–Arg	–	–	—	**27**	D-Ser–L-His–OMe	—	–	—
12	Ser–Lys	–	–	—	**28**	D-Ser–D-His–OMe	—	–	—
13	Ser–OMe	+	–	—	**29**	L-Ser–D-His–OMe	—	+	—
14	Ser–His–OMe	+	+	—	**30**	Ser–D-Phe	—	+	—
15	Boc–Ser–His	–	–	—	**31**	L-Ser–Hism	+	—	—
16	Thr–OMe	+	—	—	**32**	D-Ser–Hism	–	—	—

"+" stands for with the cleavage activity, and "–" stands for without the cleavage activity.

Ser–His, this kind of magical dipeptide, has substrate hydrolysis universality to some extent; it can hydrolyze different substrate proteins, such as BSA, GFP and CyPA. In

addition, it has the function reversibility. This means that Ser–His not only hydrolyzes peptide bonds, but also catalyzes the formation of peptide bonds. All the aforementioned features are similar with modern enzymes, such as proteases.

Besides, bioinformatics-based analysis was also performed to evaluate the evolutionary relationships between Ser–His and modern serine proteases further. The active site structures of serine proteases derived from 17 representative organisms, including prokaryotes such as *Escherichia coli*, *Synechocystis* and complex life forms such as humans and mice were analyzed, compared and summarized in our laboratory, as shown in Figure 9.9. From all of the organisms represented, the relative conservative motif Ser–[X]–His was identified as the major pattern at the catalytic sites of serine proteases, but not *Danio rerio,* which uses Ser–Lys instead. In most instances, a negatively charged Asp residue, together with Ser and His, serves as a catalytic triad for protein hydrolysis. Glu, another negatively charged amino acid, also occasionally replaces Asp at the catalytic site [3]. This result strongly supports our hypothesis that the Ser–His dipeptide is likely the origin of serine proteases.

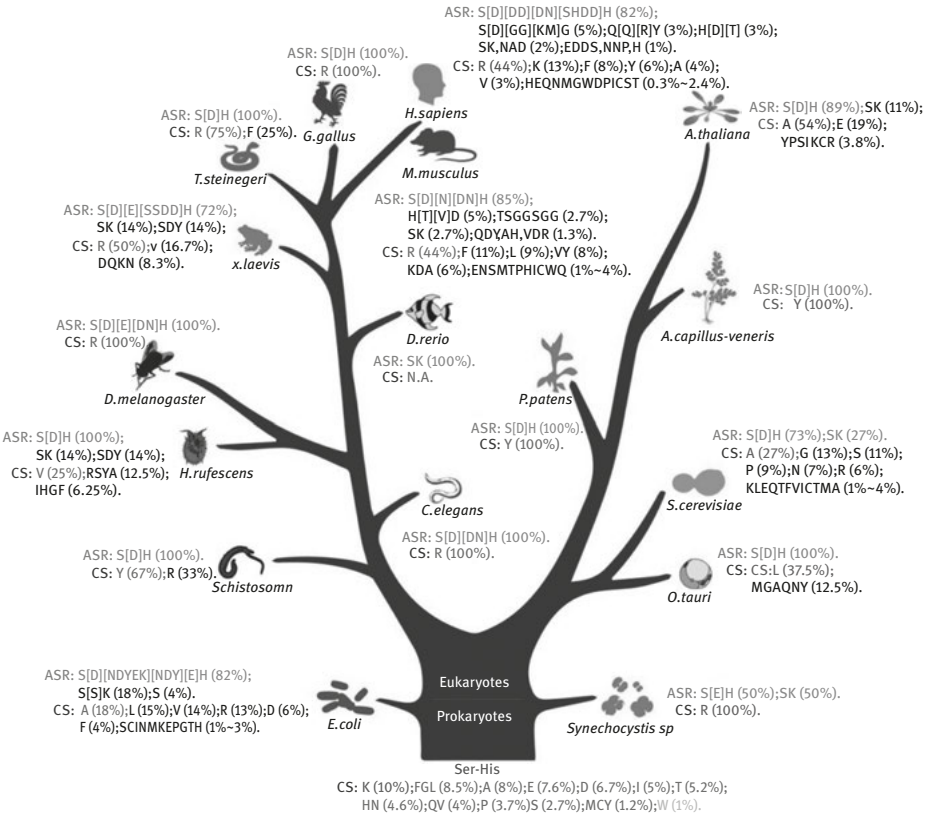


Figure 9.9: The evolutionary tree of the serine proteases. Reprinted with permission [Springer Nature].

Seventeen representative organisms were included in the analysis: *Adiantum capillus-veneris* (*A. capillus-veneris*), *Arabidopsis thaliana* (*A. thaliana*), *Caenorhabditis elegans* (*C. elegans*), *Danio rerio* (*D. rerio*), *Drosophila melanogaster* (*D. melanogaster*), *Eisenia foetida* (*E. foetida*), *Escherichia coli* (*E. coli*), *Gallus gallus* (*G. gallus*), *Haliotis rufescens* (*H. rufescens*), *Homo sapiens* (*H. sapiens*), *Mus musculus* (*M. musculus*), *Ostreococcus tauri* (*O. tauri*), *Physcomitrella patens* (*P. patens*), *Saccharomyces cerevisiae* (*S. cerevisiae*), *Schistosoma synechocystis sp, Trimeresurus stejnegeri* (*T. stejnegeri*) *and Xenopus laevis* (*X. laevis*). The percentages of amino acids at the active site residues (ASRs) and cleavage sites (CSs) were calculated subject to the known serine proteases within the selected organism. The number of serine proteases involved in the analysis is also provided beside the species name. The brackets ([]) in the ASR statistics refers to the optional amino acids, delimited by commas (,). The cleavage sites of Ser–His were determined according to mass spectrometry-based cleavage analysis of the four protein substrates undertaken in this study. The most common ASRs (green) and the most favored CSs (red) of each organism are listed first [30].

Moreover, it should be noted that Ser–His also hydrolyzes phosphodiester bond and catalyzes the formation of phosphodiester bond [31], which is just like phosphatase.

Although the cleavage activity is weak, Ser–His has the most basic function of modern enzyme. With the evolution of the structure from simple to complex of the prototype enzyme, their functions become more powerful and specific over evolutionary history. Does Ser–His exist in the prebiotic condition? People may be suspicious of this question because serine is one of the prebiotic amino acids [32, 33], but His is not. However, the synthesis of histidine in simulated prebiotic conditions has also been reported [34]. Therefore, Ser–His might be obtained in prebiotic conditions.

As the discussion above, polypeptides with one or two amino acids inserting in the middle of Ser and His residues of Ser–His possess the cleavage activities. For polypeptides prolonged from C-terminal of Ser–His, their cleavage activity still reserves. This means that Ser–His has a powerful evolutionary competence. Since the number of available combinations with catalytic activity for the evolution of polypeptide enzymes is limited due to the relatively small number of functional groups provided by the natural amino acids, the successful combination of Ser and His is the repeatedly selected result in the evolution of diverse groups of enzymes. Either on the view of molecular structural evolution or functional evolution, it is reasonable to believe that Ser–His is one potential candidate of the evolution prototype of modern enzyme.

References

[1] Steitz T A and Shulman R G. Crystallographic and NMR studies of the serine proteases [J]. *Annu. Rev. Biophys. Bioeng.*, 1982, 11: 419–444.

[2] Brady L, Brzozowski A M, Derewenda Z S, et al. A serine protease triad forms the catalytic centre of a triacylglycerol lipase [J]. *Nature*, 1990, 343(6260): 767–770.

[3] Schrag J D, Li Y G, Wu S, et al. Ser-His-Glu triad forms the catalytic site of the lipase from Geotrichum candidum [J]. *Nature*, 1991, 351(6329): 761–764.
[4] Sussman J L, Harel M, Frolow F, et al. Atomic structure of acetylcholinesterase from Torpedo californica: a prototypic acetylcholine-binding protein [J]. *Science*, 1991, 253(5022): 872–879.
[5] Anthonsen, H W, Baptista A, Drablos F, et al. Lipases and esterases: a review of their sequences, structure and evolution [J]. *Biotechnol. Annu. Rev.*, 1995, 1: 315–371.
[6] Hedstrom L. Serine protease mechanism and specificity [J]. *Chem. Rev.*, 2002, 102(12): 4501–4524.
[7] Carter P and Wells J A. Dissecting the catalytic triad of a serine protease [J]. *Nature*, 1988, 332 (6164): 564–568.
[8] Corey D R, Willett W S, Coombs G S, et al. Trypsin specificity increased through substrate-assisted catalysis [J]. *Biochemistry*, 1995, 34(36): 11521–11527.
[9] Craik C S, Roczniak S, Largman C, et al. The catalytic role of the active site aspartic acid in serine proteases [J]. *Science*, 1987, 237(4817): 909–913.
[10] Paetzel M and Dalbey R E. Catalytic hydroxyl/amine dyads within serine proteases [J]. *Trends. Biochem*. Sci., 1997, 22(1): 28–31.
[11] Li Y S, Zhao Y F, Hatfield S, et al. Dipeptide seryl-histidine and related oligopeptides cleave DNA, protein, and a carboxyl ester [J]. *Bioorg. Med. Chem.*, 2000, 8(12): 2675–2680.
[12] Wan R, Wang N, and Zhao Y F. Studies on DNA cleaved by seryl-histidine dipeptide [J]. Chem. *J. Chinese U.*, 2001, 22(4): 598–600.
[13] Wan R, Wang N, Zhao G, et al. Seryl-histidine dipeptide side chains effects on its DNA cleavage reaction [J]. *Chem. J. Chinese U.*, 2000, 21(12): 1864–1866.
[14] Reddy P R, Rao K S, and Mohan S K. Copper(II) complexes containing N,N-donor ligands and dipeptides act as hydrolytic DNA-cleavage agents [J]. *Chem. Biodivers.*, 2004, 1(6): 839–853.
[15] Santoro S W, Joyce G F, Sakthivel K, et al. RNA cleavage by a DNA enzyme with extended chemical functionality [J]. *J. Am. Chem. Soc.*, 2000, 122(11): 2433–2439.
[16] Beloglazova N G, Fabani M M, Zenkova M A, et al. Sequence-specific artificial ribonucleases. I. Bis-imidazole-containing oligonucleotide conjugates prepared using precursor-based strategy [J]. *Nucleic Acids Res.*, 2004, 32(13): 3887–3897.
[17] Ma Y, Chen X, Sun M, et al. DNA cleavage function of seryl-histidine dipeptide and its application [J], *Amino Acids*, 2008, 35(2): 251–256.
[18] Sun M, Zhu C J, Tan B, et al. Molecular modeling of diastereoisomeric aggregates of L/D ser/ histamine amide with 5 '-TpTpdC-3 ' [J], *J. Mol. Model.*, 2003, 9(2): 84–87.
[19] Zhu C J, Lu Y, Du H L, et al. A pair of new enantiomeric amides with DNA-cleaving function [J]. *Chem. J. Chinese U.*, 2004, 25(6): 1065–1068.
[20] Sun M, Ma Y, Ji S H, et al. Molecular modeling on DNA cleavage activity of seryl-histidine and related dipeptide [J]. *Bioorg. Med. Chem. Lett.*, 2004, 14(14): 3711–3714.
[21] Liu X H. Sun M, Cheng C M, et al. Theoretical study on aggregates of seryl-histidine and histidyl-serine dipeptide with 5 '-TpTpdC-3 ' [J]. *Orig. Life Evol. Biosph.*, 2006, 36(3): 263–265.
[22] Zhong R G, Zhao L J, and Zhao Y F. A QM-MM study on DNA cleavage by seryl-histidine dipeptide [J]. *Acta Chimica Sinica*, 2004, 62(24): 2444–2446.
[23] Chen J, Wan R, Liu H, et al. Cleavage of BSA by a dipeptide seryl-histidine [J]. *Lett. Pept. Sci.*, 2000, **7**(6): 325–329.
[24] Du H L, Wang Y T, Yang L F, et al. Appraisal of green fluorescent protein as a model substrate for seryl-histidine dipeptide cleaving agent [J]. *Lett. Pept. Sci.*, 2002, 9(1): 5–10.
[25] Liu Y, Shi Y H, Liu X X, et al. Evaluation of non-covalent interaction between Seryl-Histidine dipeptide and cyclophilin A using NMR and molecular modeling [J]. *Sci. China Chem.*, 2010, 53(9): 1987–1993.

[26] Chen J, Wan R, Liu H, et al. Studies on the cleavage of bovine serum albumin by Ser-His [J]. *Chem. J. Chinese U.*, 2001, 22(8): 1349–1351.
[27] Jakubke H D, Kuhl P, Könecke A. Basic principles of proteasecatalyzed peptide bond formation [J]. *Angew Chem., Int. Ed. Engl.*, 1985, 24: 85–93.
[28] Gorlero M, Wieczorek R, Adamala K, et al. Ser-His catalyses the formation of peptides and PNAs [J]. *Febs Lett.*, 2009, 583(1): 153–156.
[29] Adamala K and Szostak J W. Competition between model protocells driven by an encapsulated catalyst [J]. *Nat. Chem.*, 2013, 5: 634–634.
[30] Liu Y, Li YB, Gao X, et al. Evolutionary relationships between seryl histidine dipeptide and modern serine proteases from the analysis based on mass spectrometry and bioinformatics [J]. *Amino Acids*. 2018, 50: 69-77.
[31] Wieczorek R, Dorr M, Chotera A, et al. Formation of RNA phosphodiester bond by histidine-containing dipeptides [J]. *Chembiochem.*, 2013, 14: 217–223.
[32] Zaia, D.A.M.; Zaia, C.T.B.V.; Santana, H.D. Which amino acids should be used in prebiotic chemistry studies [J]? *Orig. Life Evol. Biosph.* 2008, 38, 469–488.
[33] Longo, L.M.; Blaber, M. Protein design at the interface of the pre-biotic and biotic worlds [J]. *Arch. Biochem. Biophys.*2012, 526, 16–21.
[34] Shen C, Yang L, Miller S L, et al. Prebiotic synthesis of histidine [J]. *J. Mol. Evol.*, 1990, 31: 167–174.

10 The interaction between ATP and amino acids

The adenosine triphosphate (ATP)-binding protein is the oldest known protein, which was found about 3.7 billion years ago. We know that ATP plays an important role in life; therefore, bioinformatics experts believe that ATP, nicotinamide adenine dinucleotide (NAD), nicotinamide adenine dinucleotide phosphate (NADP) and other small molecules induce the production and evolution of important protein families. The molecular interaction between protein and ATP determines the biological functions of all proteins. As amino acids (AAs) are the building block of the proteins, the study of the interaction between ATP and amino acids is a subject of great importance.

10.1 The study of the origin and evolution of protein based on small molecule

10.1.1 ATP – a living fossil at the molecular level

Macromolecules (proteins, DNA, RNA, etc.) and small molecules (coenzyme, metabolic products, metal ions, etc.) are significant for life. Biological macromolecules are the main research subjects for traditional evolutionary biology, because their rich sequence and structure are suitable for studying the evolutionary mechanism of life. Small molecules change much more slowly than macromolecules. Most small molecules are very conservative, and the differences between the species are small. For example, both humans and bacteria use ATP as their main source of energy. It can be seen that the information obtained from small biological molecules is far less than that from large molecules, and therefore small biological molecules have been neglected by evolutionary biology for a long time. However, everything must be treated dialectically. The conservation of small molecules makes them a living fossil at the molecular level (called molecular fossils), which is used to trace the mystery of origin and evolution of life.

The mechanism of origin of life and evolution is one of the core topics in natural science research. The research of the origin of modern life originated in 1953, when Miller successfully got AAs using gases such as methane, hydrogen, ammonia and water vapor. After that, the research was basically performed with the prebiotic synthesis of a number of small molecules that composed life. However, the progress of research for origin of biological macromolecules, especially protein, was relatively slow, because it is very difficult to trace the structure and function of the original proteins that existed billions of years ago. Only in recent years, there has been progress in research in this field. With the rapid development of various kinds of "omics," a lot of information about the protein structure and function has been found. From this, the researcher can infer the original protein characteristics using bioinformatics.

https://doi.org/10.1515/9783110562552-010

10.1.2 ATP-binding protein – the oldest protein

Professor Caetano-Anollés found that, using phylogenomic analysis, c.37, c.1, c.2, c.23, c.26 fold were the most ancient architectures of proteins [1]. Furthermore, they concluded that these folds appeared in purine metabolism, pyrimidine metabolism and porphyrin and chlorophyll metabolism [2]. Professor Trifonov also proposed c.37 folding type appeared most early, after they analyzed the sequence and structure characteristics of a series of proteins, and they also found the original proteins needed the help of ATP as a cofactor [3]. It was also found that most of the ancient proteins function with the help of a phosphate group cofactor such as ATP, NAD and NADP [4], which serves as evidence for Professor Trifonov's opinion. The aforementioned research methods are basically independent of each other, and so it can be concluded that the common results are credible, which the folding type of most primitive protein is c.37, C.1, C.2, and ATP, NAD, NADP act as a cofactor for protein to complete their normal function. However, there are a series of more challenging problems. Why c.37, C.1 and C.2 are first folding type? Why these proteins select ATP, NAD and NADP as cofactors? Whether we can use the method to prove folded protein structure of c.37, C.1 and C.2 type is the most primitive?

It is noted that the most ancient protein contains a common coenzyme or cofactor. Cofactors originated earlier than proteins. Thus, we propose to use conservative small molecules as "molecular fossils" for the research of origin of protein. By analyzing the distribution patterns in the protein space of 2,000 kinds of small molecules in the sc-PDB database, the scientists have found that a few small molecules, such as ATP, NAD/P, FAD and FMN, can be combined with a variety of proteins, while most of the small molecules can be combined with only a few or even a single protein. It exhibits a power-law distribution (Figure 10.1). This distribution could be interpreted by the preferential attachment principle. The distribution pattern of small molecule in protein conformation gives the record of the evolution of small molecules and protein binding. The more widely

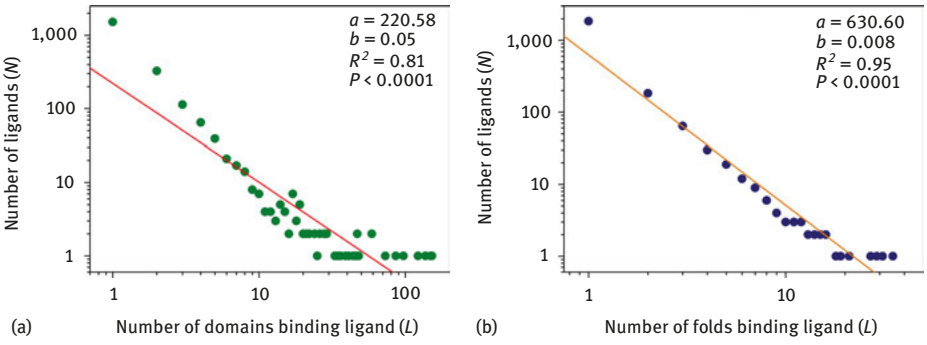


Figure 10.1: A power-law distribution of the small molecules in protein conformation. The number of small molecules(N)and the number of the binding proteins(L)are in accordance with formula $N = aL^{-b}$.

distributed small molecules in protein conformation, the earlier they bind to proteins. Hence, it can be inferred that ATP, NAD/P, FAD and FMN are the first small-molecule ligands with primitive protein, and the fold type of the conjugated protein belongs to c.37, C.2, C.3, c.23 and c.26 [4], which is the same as suggested by Professor Caetano-Anollés. This consistency is not accidental. The best explanation is the interaction between protein and small molecules has a connection with the origin of protein. It is put forward about "the origin of small molecule induction or selection model for the protein origin study", which means original proteins come from the selection of small molecule ligands from random peptide pool and ATP chooses the ancient protein fold type as c.37 fold [5].

The model has many theoretical and experimental evidence. First, through a large number of experimental studies, it has been found that small molecule ligands do have effect on the structure formation of protein [6]. Second, in 2001, Professor Szostak in Harvard University used the mRNA display technology to fish a random peptide that can be combined with ATP from a random peptide library with a storage capacity of 10^{12}/mL [7]. He found that the structure of the protein he had fished was very similar to the c.37 fold type [8, 9]. Hydrolysis of ATP was the function of that protein [10]. It is fully consistent with the function of ATP phosphate hydrolase [11], which is the theoretically predicted most ancient protein. A series of study show that small molecules do play an important role in the origin of protein. Thereafter, Professor Tawfik published an article in the *Science* [12], and he also proposed that the origin of proteins was related to the selection of small molecules.

The conserved structure of small molecule not only helps to trace the origin of protein, but also helps to interpret large-scale evolution of proteins. It is well known that the evolution of protein structure is very slow, which means that labeled protein structures with time scale can be used as a molecular clock. Many small biological molecules, such as porphyrin, steroid, flavonoid, are highly conserved with clear geological record. At the same time, their synthetase has a unique structure. Therefore, these small molecules can establish a relationship between protein structure and geological age, which can be used to determine the evolutionary chronologies of protein structure. By the cooperation of Professor Caetano-Anollés in Illinois University, professor Zhang group established the evolution sequence of 1,030 protein folding types and 1,730 super families. They also got their time scale by geological age of small molecules, and established the molecular clock based on protein structure shown in Figure 10.2 [13].

10.2 The structure and the interaction with AAs of ATP

10.2.1 The structure of ATP

As shown in Figure 10.3, ATP is the most widely distributed and the most important nucleoside in organisms. It not only functions as an energy carrier, but also regulates

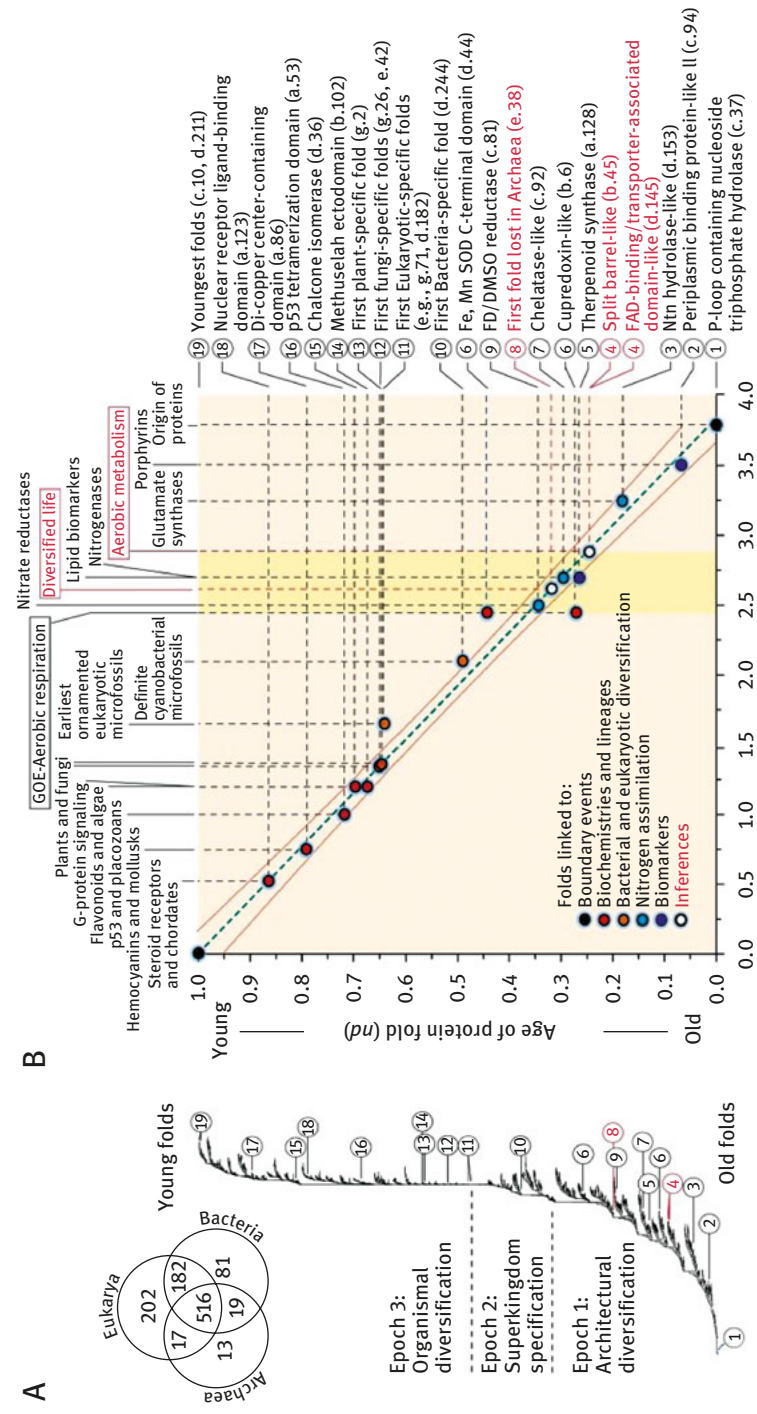


Figure 10.2: (A) Phylogenomic tree of folds. (B) The molecular clock based on protein fold type and the recorded important evolutionary events.

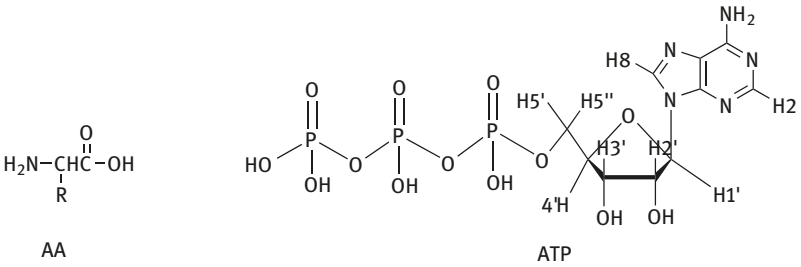


Figure 10.3: The structure of AAs and ATP.

complex biochemical processes via phosphorylation of proteins. Therefore, it is of great importance to study the molecular recognition of ATP by enzyme, the special proteins in vital movement [14]. An enzyme for ATP can be recognized by the intermolecular interaction between ATP and AA residues on the enzyme, such as intermolecular hydrogen bond, Van der Waals force, hydrophobic interaction, π–π interactions of aromatic systems, and so on [15-17]. As shown in Figure 10.3 and Table 10.1, AAs are the building block of the proteins and peptides, and can be used as a minimodel of enzyme to investigate its interaction with ATP.

Table 10.1: The information about 20 AAs

Name*	Abbreviation	Molecular weight	R
Glycine	Gly	75.1	–H
Alanine	Ala	89.1	$-CH_3$
Valine	Val	117.1	$-CH(CH_3)_2$
Leucine	Leu	131.2	$-CH_2CH(CH_3)_2$
Isoleucine	Ile	131.2	$-CH(CH_3)\ CH_2$
Serine	Ser	105.1	$-CH_2OH$
Threonine	Thr	119.1	$-CHOHCH_3$
Cysteine	Cys	240.3	$-CH_2SH$
Methionine	Met	149.2	$-\ (CH2)_2SCH_3$
Proline	Pro	115.1	$-\ (CH_2)_3-$
Aspartic acid	Asp	133.1	$-CH_2COOH$
Asparaginate	Asn	132.1	$-CH_2CONH_2$
Glutamic acid	Glu	147.1	$-CH_2CH_2COOH$
Glutamine	Gln	146.1	$-CH_2CH_2CONH_2$
Phenylalanine	Phe	165.2	$-C_6H_5$
Tyrosine	Tyr	181.2	$-C_6H_4OH$
Tryptophan	Trp	204.1	$-CH_2C_8H_6N$
Lysine	Lys	146.2	– (CH2)4NH2
Arginine	Arg	174.1	$-\ (CH_2)_3CH(NH)NH_2$
Histidine	His	165.2	$-CH_2C_3H_3N_2$

* For convenience, the mentioned AAs are abbreviated.

Electrospray ionization (ESI) technique is sufficiently soft to allow the existing of weak non-covalent interaction in the ionization process of noncovalent complexes, and then detection of these complex can be completed in the mass spectrum. On the basis of the signal of the complex, it is easy to derive the stoichiometric information of the complex [18, 19]. In addition, ESI-MS (mass spectrometry) technology has good specificity, high sensitivity, fast speed. In recent years, ESI-MS technique has become a useful tool to assess noncovalent interaction of biomolecules [20-22]. Now the use of ESI-MS is widespread, such as to determine binding constants that can be used to know the strength of weak interaction, to study the selection of intermolecular noncovalent interaction and to study binding sites of complexes [23-25]. In our laboratory, ESI-MS technique was used to study the noncovalent interaction between ATP and 20 AAs to examine molecular recognition of ATP by enzyme. The key to success in the study depends on careful understanding and manipulation of ESI source parameters. The complex ions would disappear by increasing the parameter of capillary exit of MS, and so the capillary parameter was tuned to compare the affinity of ATP with different AAs [26]. The noncovalent interactions between ATP and aromatic AA were also investigated and binding constants, the quantitative results of weak interaction, were obtained by fluorescence spectrometry. Aromatic ring stacking effect of aromatic AAs and ATP was also studied by nuclear magnetic resonance (NMR).

In recent years, computational chemistry has developed rapidly with the advances of computer technology, and has gradually become a routine research tool in the field of chemistry and biological research. In our laboratory, on the basis of molecular mechanics method, molecular models of ATP and AAs with different functional groups were established and optimized using Sybyl 7.1 software. Using these models, the weak interaction between ATP and AA was studied. Noncovalent complexes models of AA and ATP were established by molecular docking and were optimized using the Tripos force field. They can be used to analyze interaction strength, weak interaction type and the effect of AA structure on the weak interaction.

10.2.2 Study on the weak interaction between AA and ATP by MS

The noncovalent interactions between ATP and 19 kinds of AAs were investigated by ESI-MS/MS. The noncovalent complexes of ATP and AA were observed (except those of ATP and Gly, Ala, Val) by mass spectra, as listed in Table 10.2. For the samples of ATP mixed with Ile, ESI-MS showed that more peaks were observed in the spectra, for example, $[ATP\text{-}3Na+Ile]^+$ at m/z 705, $[ATP\text{-}4Na+Ile]^+$ at m/z 727 and $[ATP\text{-}5Na+Ile]^+$ at m/z 749. For the samples of ATP mixed with Phe, such complex ions were observed as $[ATP\text{-}3Na+Phe]^+$ at m/z 739, $[ATP\text{-}4Na+Phe]^+$ at m/z 761 and $[ATP\text{-}5Na+Phe]^+$ at m/z 783. The complex ions of ATP and other AAs were also observed by mass spectra. On the basis of the molecular weight of complexes, the stoichiometric ratio of the two should be 1:1.

Table 10.2: The noncovalent complexes of ATP and AA detected by ESI-MS

Name of AA	Complex ions (*m/z*)			
	$[ATP\text{-}2Na+AA]^+$	$[ATP\text{-}3Na+AA]^+$	$[ATP\text{-}4Na+AA]^+$	$[ATP\text{-}5Na+AA]^+$
Ile	/	705	727	749
Leu	/	705	727	749
Ser	/	679	701	723
Thr	/	693	715	737
Met	/	723	745	767
Pro	/	689	711	733
Asp	/	707	729	751
Asn	/	706	728	750
Glu	/	721	743	765
Gln	/	720	742	764
Phe	/	739	761	783
Tyr	/	755	777	799
Trp	/	778	800	822
Lys	698	720	/	/
Arg	726	748	/	/
His	707	729	/	/

"/": the noncovalent complex ions were not observed.

To get further structural information of the complex, multistage MS was carried out. In MS, the first step was to choose the target ion. Only when the weak interaction was strong, these complexes were stable enough to be isolated and fragmented. Then fragment ions of the complex ions could be obtained. On the other hand, if molecular interaction is weak, the complex ions would dissapear in the isolated process. Only MS/MS spectra of adduct ions of ATP with Phe were obtained as shown in Figure 10.4. They all lost a neutral molecule with the mass 165 Da, which is Phe. The fragment ions were the ions of ATP. This indicates that the weak interactions were destructed in the fragmentation process. This result also shows that Phe and ATP were combined together with the ratio of 1:1.

The stability of the complexes depends on the cone voltage of MS spectrometer. The weak interaction will be destroyed and the complex ions will disappear by increasing the value of the cone voltage. Thus, when the complex ions disappear, the cone voltage can be used to compare the strength of weak interaction [26]. The data for ATP and AA are listed in Table 10.3. According to the data, the affinity sequence of AA for ATP is: Phe > AAs with OH, COOH, SH, $-CONH_2$, NH_2 on the side chain > AAs with an alkyl substituent on the side chain. In combination with the structure of AAs, the factors that influence the interaction are analyzed. The noncovalent complexes between ATP and Ile, Leu or Pro were observed, while no interaction between ATP and Gly, Ala or Val was found. This may be due to the enhancement of hydrophobic effect with the increase of alkyl chain on side chain. The side chain played an important role in the noncovalent

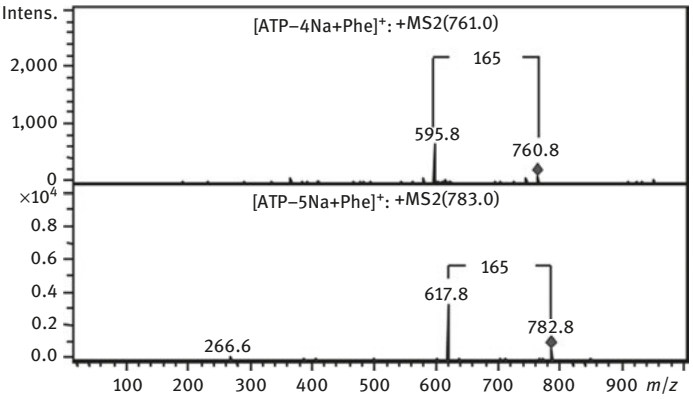


Figure 10.4: MS/MS spectra of noncovalent complexes of Phe and ATP.

interaction between ATP and AAs. On the side chain of Ser, Thr, Met, Asp, Asn, Glu, Gln, Arg, Lys and His, the functional groups were hydrogen bond acceptor and donor. They can form a strong hydrogen bond with ATP, which would enhance affinity. According to the mass spectra, the affinity sequence was as follows: Met, Ser, Thr <Asp, Glu, Lys <Asn, Gln, Arg, His. Considering the structure of the side chain group, the affinity sequence was: R=C–NH_2>–RCOOH, –R–NH_2>, –RSH, –ROH. The complex ions between Phe and ATP were most stable because of the strong hydrophobic interaction and π–π interactions with the purine ring of ATP induced by the phenyl group on the side chain of Phe. From the aforementioned discussion, it could be summed up that the contribution of the side chain groups to weak interactions from low to high was: –R–NH_2> –RSH, phenyl >R=C–NH_2>–RCOOH, –ROH > long alkyl > short alkyl.

Table 10.3: The data of cone voltage when the noncovalent complex ions disappeared

Name of AA	Cone voltage (volt)	Name of AA	Cone voltage (volt)
Gly*	/	Asn	180.3
Ala*	/	Glu	170.4
Val*	/	Gln	191.0
Ile	132.7	Phe	210.4
Leu	139.2	Tyr**	/
Ser	156.0	Trp**	/
Thr	148.3	Lys	169.5
Met	154.8	Arg	185.3
Pro	140.6	His	185.3
Asp	164.7		

* The noncovalent complex ions between ATP and Gly, Ala or Val were not observed.
** The data of cone voltage for the noncovalent complex ions between ATP and Tyr or Trp cannot be obtained because of the interference of other ions.

10.2.3 Study on the weak interaction between AA and ATP by fluorescence spectrometry

The AAs containing aromatic rings have strong fluorescence. Fluorescence quenching occurs when ATP is added to the solution of Phe and Trp [27,28]. Therefore, the binding constant for ATP and Trp or Phe could be obtained using fluorescence spectrometry. Fluorescence quenching of Phe and Trp in the presence of ATP is shown in Figures 10.5 and 10.6. MS results showed that complexes of ATP and AAs were at the ratio of 1:1. Thus, the dissociation constants were calculated according to formula (1) [27], and the binding constants calculated according to formula (2):

$$I = [I_0 + I_\infty[L]/K_d]/(1+[L]/K_d) \quad (1)$$

$$K = 1/K_d \quad (2)$$

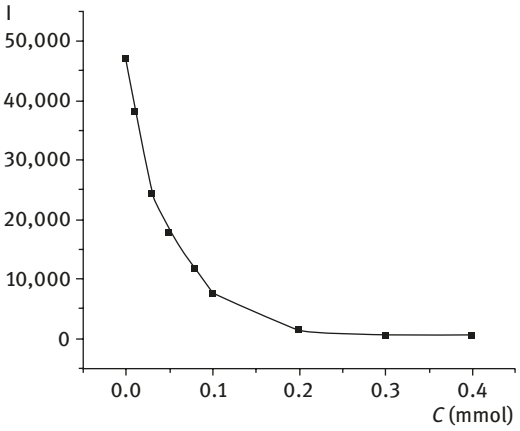


Figure 10.5: Fluorescence quenching of Phe in the presence of ATP.

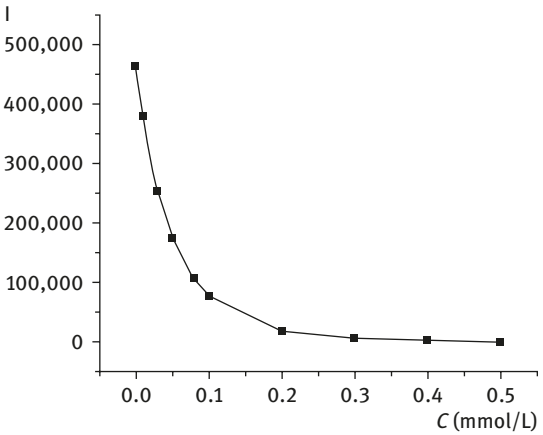


Figure 10.6: Fluorescence quenching of Trp in the presence of ATP.

where I_0 is the fluorescence intensity of Phe or Trp without ATP. I_∞ is fluorescence intensity when ATP and the AA combine completely. [L] is the concentration of ATP. K_d represents dissociation constants and K stands for binding constants.

As listed in Table 10.4, the binding constants of ATP and Phe are 38.47 mmol^{-1}, while that of Trp is 39.23 mmol^{-1}. Thus, the affinity of Trp with ATP was stronger than that of Phe.

Table 10.4: Dissociation constants and binding constants

Name of AA	excitation wavelength (nm)	K_d (mmol)	K (mmol^{-1})
Phe	258	0.026	38.47
Trp	280	0.025	39.23

10.2.4 Study of the weak interaction between AA and ATP by NMR

Both ATP and aromatic AAs have aromatic ring. It has been reported that when the proteins recognize ATP, the aryl of aromatic AAs and the adenine ring of ATP have π–π stacking interactions [29]. We have studied this phenomenon using NMR technology. The solubility of Tyr is poor and NMR analysis cannot be performed. As listed in Table 10.5, chemical shifts of ATP proton mixed with Phe at different concentration ratios are different. In the presence of Phe, H2 and H8 on the purine ring moiety of ATP shift to a higher field. The higher the concentration of Phe, the higher field the protons shift to. This is because of the π–π stacking interaction between Phe and the purine phenyl ring of ATP. H2 and H8 are affected by magnetic anisotropy effect of the electronic circulation in phenyl moiety of Phe, which makes chemical shifts of H2 and H8 change.

Table 10.5: Chemical shift (ppm) of ATP proton in the presence of Phe

Name (concentration ratio)	δH8	δH2	δH1″
ATP	8.601	8.437	6.123
ATP+Phe (1:1)	8.592	8.398	6.108
Δδ	−0.009	−0.039	−0.015
ATP+Phe (1:2)	8.58	8.361	6.092
Δδ	−0.021	−0.076	−0.031

Because of the poor solubility of Trp in the water, only the mixed solution of ATP and Trp with the concentration ratio of 1:1 was studied by NMR. As listed in Table 10.6, influenced by magnetic anisotropy effect of the electronic circulation of Trp, H2, H8 and H1′ on ATP shifted to higher field (0.275, 0.101 and 0.108 ppm, respectively).

Table 10.6: Chemical shift (ppm) of ATP proton in the presence of Trp

Name	δH8	δH2	δH1′
ATP	8.595	8.432	6.116
ATP+Trp	8.494	8.157	6.008
Δδ	−0.101	−0.275	−0.108

Hence, it can be concluded that the π–π stacking interaction between Trp and ATP is stronger than that between Phe and ATP.

10.2.5 Study of the weak interaction between AA and ATP by theoretical calculation

We selected several AAs with different functional groups for molecular dynamic simulation to study their interaction with ATP. The structures of noncovalent complexes with lowest binding energy were obtained as shown in Figures 10.7, 10.8, 10.9, 10.10 and 10.11 using the DOCK method in Sybyl 7.1. It had been found that hydrogen bonds can be formed between AAs and ATP, as shown in Table 10.7. The hydrogen bonds in the complex were determined according to criteria that were used by McDonald and Thornton, with a maximum donor-to-acceptor distance of 3.5 Å, a maximum hydrogen atom to acceptor distance of 2.5 Å and angles of at least 90° at both the hydrogen atom and the acceptor. The C=O and N–H unit on the main chain of Ala formed two strong hydrogen bonds, with the triphosphate moiety and NH_2 on the base moiety of ATP. The C=O on the main chain of Phe formed a hydrogen bond with NH_2 on the base moiety of ATP. In the complex of ATP and Asp or Ser, there were three hydrogen bonds, while there were four hydrogen bonds in the complex of ATP and Asn. These intermolecular hydrogen bonds make the complex stable.

The binding energies were also calculated using the Tripos force field in Sybyl 7.1, as listed in Table 10.8. The binding energy of Asn and ATP is –17.702 kJ/mol, which is lowest. The binding energy of Asp, Ser and Phe is low. For Ala, the binding energy is –6.555 kJ/mol, which is so high that the complex ions cannot be observed by the mass spectra. When the binding energy is low, the noncovalent complex is more stable and the intermolecular interaction is stronger. Therefore, the affinity sequence of the

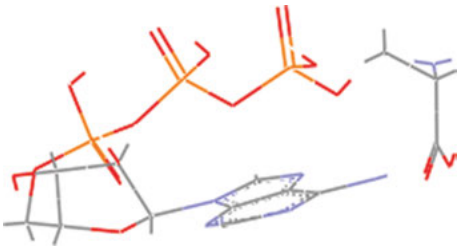


Figure 10.7: The noncovalent complex model of ATP and Ala.

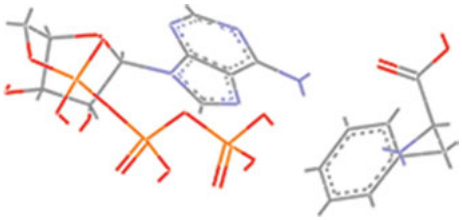


Figure 10.8: The noncovalent complex model of ATP and Phe.

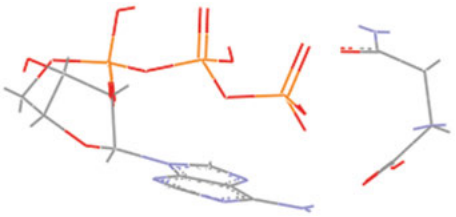


Figure 10.9: The noncovalent complex model of ATP and Asn.

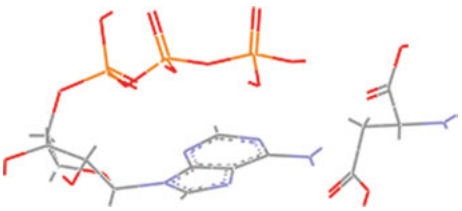


Figure 10.10: The noncovalent complex of ATP and Asp.

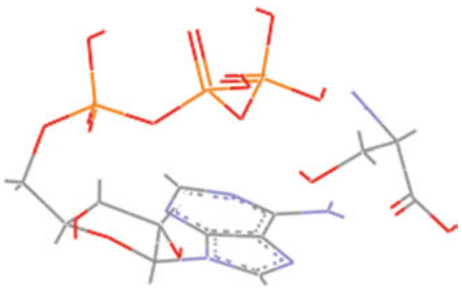


Figure 10.11: The noncovalent complex model of ATP and Ser.

AAs for ATP is obtained: Asn> Asp > Ser > Phe > Ala, which is favorably consistent with the conclusion acquired by ESI-MS except for Phe. The π–π stacking interaction plays a significant role in the noncovalent interaction between Phe and ATP, which our NMR results have confirmed. Because we are using molecular mechanics that is not primarily concerned with the electronic circulation of electrons, this simulation method may not be suitable for Phe and ATP.

Table 10.7: The hydrogen bond in the noncovalent complexes of ATP and AA

Complex name	The information of the hydrogen bond				
	L–AA	ATP	Length	Angle	X–Y
ATP–Ala	C=O*	N–H***	1.640	163.04	2.644
	N–H*	P(γ) –O	1.856	160.39	2.851
ATP–Phe	C=O*	N–H***	1.675	135.18	2.515
ATP–Asn	C=O*	N–H***	1.643	165.73	2.658
	N–H*	P(γ) –O	2.571	132.66	3.359
	C=O**	P(γ) –Ona	2.146	120.56	2.755
	N–H**	P(γ)=O	1.605	130.47	2.395
ATP–asp	C=O*	N–H***	1.653	144.23	2.567
	C=O**	N–H***	1.623	169.75	2.649
	C=O**	P(γ) –Ona	1.838	136.34	2.612
ATP–Ser	C=O(1)	N–H***	1.640	168.90	2.660
	N–H*	P(γ) –Ona	1.903	152.82	2.684
	C–O**	N–H***	2.013	139.25	2.708

*The group on the main chain of AAs. **The group on the side chain of AAs. ***NH_2 on the adenine moiety of ATP.

Table 10.8: ATP – the binding energy of noncovalent complex(kJ/mol)

AAs	Energy			
A	$E_{A\text{-}ATP}$	E_A	E_{ATP}	ΔE
Ala	–54.278	1.931	–49.654	–6.555
Phe	–58.994	1.841	–49.654	–11.181
Asp	–75.174	–8.199	–49.654	–17.321
Asn	–77.262	–9.906	–49.654	–17.702
Ser	–57.950	7.264	–49.654	–15.559

10.2.6 The interaction between AA and ATP

The noncovalent interactions between ATP and 19 kinds of AAs were investigated by ESI-MS/MS, fluorescence spectrometry, NMR and theoretical calculation. The noncovalent complexes of ATP and AA were observed except for those of ATP and Gly, Ala, Val in mass spectra. The influence of the group on the side chain of AAs on the noncovalent interactions was analyzed by comparing the stability of the complexes. The noncovalent interactions between ATP and Phe or Trp were also investigated by fluorescence spectrometry and NMR. The affinity sequence of AA for ATP was obtained as follows: Trp > Phe > AAs with R=C–NH_2 on the side chain > AAs with –RCOOH, –R–NH_2 on the side chain > AAs with –RSH, ROH on the side chain > AAs with long alkyl substituent > AAs with short alkyl substituent. The

result of theoretical calculation was favorably consistent with that of experiments except for Phe. By theoretical calculation, the intermolecular hydrogen bonds were confirmed to stabilize the noncovalent complexes.

The noncovalent interactions between ATP and AAs play an important role in the molecule recognition of ATP by proteins. The conclusion will throw light on the prediction of ATP-binding site and the study of molecule recognition mechanism of ATP. Like biological macromolecules, small biological molecules also contain mysteries of life, even the mysteries of changes of the Earth's environment. As long as appropriate bioinformatics methods and various spectral techniques are applied, all the mysteries might be effectively interpreted.

References

[1] Caetano-Anollés G, Caetano-Anollés D, An evolutionarily structured universe of protein architecture [J]. *Genome Res.,* 2003, 13:7 1563–1571.
[2] Caetano-Anollés G, Kim H S, Mittenthal J E, The origin of modern metabolic networks inferred from phylogenomic analysis of protein architecture [J]. *Proc. Natl. Acad. Sci. USA* 2007, 104:22 9358–9363.
[3] Trifonov E N, Early molecular evolution [J]. *Israel J. Ecol. Evol.,* 2006, 52:2 375–387.
[4] Ma B G, Chen L, Ji H F, et al., Characters of very ancient proteins [J]. *Biochem. Biophys. Res. Commun.,* 2008, 366:3 607–611.
[5] Ji H F, Kong D X, Shen L, et al., Distribution patterns of small-molecule ligands in the protein universe and implications for origin of life and drug discovery [J]. *Genome Biol.,* 2007, 8:8 R176.
[6] Cao Y, Li H, Dynamics of protein folding and cofactor binding monitored by single-molecule force spectroscopy [J]. *Biophys. J.,* 2011, 101:8 2009–2017.
[7] Keefe A D, Szostak J W, Functional proteins from a random-sequence library [J]. *Nature,* 2001, 410:6829 715–718.
[8] Surdo P L, Walsh M A, Sollazzo M, A novel ADP- and zinc-binding fold from function-directed *in vitro* evolution [J]. *Nat. Struct. Mol. Biol.,* 2004, 11:4 382–383.
[9] Caetano-Anollés G, Kim K M, Caetano-Anollés D. The phylogenomic roots of modern biochemistry: origins of proteins, cofactors and protein biosynthesis [J]. *J. Mol. Evol.* 2012, 74:1–2 1–34.
[10] Simmons C R, Stomel J M, McConnell M D, et al, A synthetic protein selected for ligand binding affinity mediates ATP hydrolysis [J]. *ACS Chem. Biol.* 2009, 4:8 649–658.
[11] Kim K M, Qin T, Jiang Y Y, et al, Caetano-Anollés D, Zhang H Y, Caetano-Anollés G, Protein domain structure uncovers the origin of aerobic metabolism and the rise of planetary oxygen [J]. *Structure,* 2012, 20:1 67–76.
[12] Tokuriki N, Tawfik D, Protein dynamism and evolvability [J]. *Science,* 2009, 324:5924 203–207.
[13] Wang M, Jiang Y Y, Kim K M, Qu, G., Ji, H. F., Mittenthal J E, Zhang H Y, Caetano-Anollés G, A universal molecular clock of protein folds and its power in tracing the early history of aerobic metabolism and planet oxygenation [J]. *Mol. Biol. Evol.,* 2011, 28(1):567–582.
[14] Chene P. ATPases as drug targets: learning from their structure [J]. *Nat. Rev. Drug Discovery*, 2002, 1:9 665–673.
[15] Xiao B, Heath R, Saiu P, et al, Structural basis for AMP binding to mammalian AMP- activated protein kinase [J]. *Nature*, 2007, 449:7161 496–501.

[16] Wong L, Jennings P A, Adams J A. Communication pathways between the nucleotide pocket and distal regulatory sites in protein kniases [J]. *Acc. Chem. Res.*, 2004, 37:5 304–311.
[17] Gradia S, Dubramanian D, Wilson T, et al,hMSH2 and hMSH6 play distinct roles in mismatch binding and contribute differently to the ATPase activity of hMutS [J]. *Molecular Cell*, 1999, 3: 255–261.
[18] Jørgensena T J D, Delforgeb D, Remacleb J, et al,Collision-induced dissociation of noncovalent complexes between vancomycin antibiotics and peptide ligand stereoisomers: evidence for molecular recognition in the gas phase [J]. *Int. J Mass Spectrom.*, 1999, 188:1–2 63–85.
[19] Friess S D, Daniel J M, Rudolf Hartmann, et al, Mass spectrometric noncovalent probing of amino acids in peptides and proteins [J]. *Int J Mass Spectrom.*, 2002, 219:1269–281.
[20] Wang W, Donini O, Reyes C M, et al, BIOMOLECULAR SIMULATIONS: Recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions [J]. *Ann. Rev. Biophys. Biomolecular Struct.*, 2001, 30: 211–243.
[21] Smith R D, Bruce J E, Wu Q, et al, New mass spectrometric methods for the study of noncovalent associations of biopolymers [J]. *Chem. Soc. Rev.,* 1997, 26:3 191–202.
[22] Fenn J B, Mann M, Meng C K, et al, Electrospray Ionization for Mass Spectrometry of Large Biomolecules [J]. *Science*, 1989,246:4926 64–71.
[23] Gao M Q, Zhang S Q, Song F G, et al,Studies on the non-covalent complexes between oleano-lic acid and cyclodextrins using electrospray ionization tandem mass spectrometry [J]. *J. Mass Spectrom.*, 2003, 38:7 723–731.
[24] Gabelica V, Galic N, Rosu F, et al, Influence of response factors on determining equilibrium association constants of non-covalent complexes by electrospray ionization mass spectrometry [J]. *J. Mass Spectrom.*, 2003, 38:5 491–501.
[25] Kempen E, Brodbelt J, A method for the determination of binding constants by electrospray ionization mass spectrometry [J]. *Anal. Chem.*, 2000, 72:21 5411–5416.
[26] PramanikBN,BartnerPL,MirzaUA, et al,Electrospray ionization mass spectrometry for the study of non-covalent complexes: an emerging technology [J]. *J. Mass Spectrom.*1998, 33:10 911–920
[27] Butterfield S M, Sweeney M M, Waters M L. The recognition of nucleotides with model β-hairpin receptors: Investigation of critical contacts and nucleotide selectivity [J]. *J. Org. Chem.*, 2005, 70:4 1105–1114.
[28] Hiroasa Y, Roche T E. Critical role of specific ions for ligand-induced changes regulating pyruvate dehydrogenase kinase isoform [J]. *Biochem.*, 2008, 47:8 2298–2311.
[29] Cochran A G, Skelton N J, Starovasnik M A, Tryptophan zippers: Stable, monomeric β-hairpins. Proc [J]. *Nat. Acad. Sci. USA*, 2001, 98:10 5578–5583.

11 Marine and the origin of life

Under the harsh environment of early Earth, seawater can effectively prevent UV damage. Submarine hydrothermal vent was supposed to be "the cradle of life." Submarine volcanoes produced polyphosphate, which catalyzed the formation of living matters. Phosphates can be considered as a key for exploring life.

11.1 The Origin of Life

The origin of life is the process of studying the evolution of nonliving matter in the primitive life of the primitive Earth and extraterrestrial planets. The Earth was born 4 billion 500 million years ago. When and where did life on Earth originate? Furthermore, what caused the formation of life? This is a major problem that modern natural science has not yet completely solved, which is also the focus of attention and debate at present. For this problem, many well-known international scientists have put forward a variety of speculations and hypotheses. Among them there are also a lot of controversies. So far there are two main theories, one is "meteoritic origin" [1], and the other is "the origin of living systems on the earth" [2]. The extreme environment, such as lightning, ultraviolet ray, volcano and deep sea hot spring, as well as the active crustal movement on the primitive Earth play a key role in the formation of primitive life. Volcanic eruptions can produce water-soluble phosphates, which are important components of the original biological materials of RNA; "black chimney" of submarines was supposed to be the place of formatting organic matters under a high temperature and deep tremendous pressure.

According to the basic concept of the origin of life, modern biological functional macromolecules originated from the original, simple small chemicals that go through the evolution of structure and function [3]. It is the origin of chemical stage in the process of the origin of life. In this chemical origin stage, inorganic small molecules generated organic matters under natural conditions. In 1953, American chemists Harold Yuri and Stanley Miller performed the famous Miller-Yuri experiment [4]. They set up a sealing system to simulate the Earth's early atmospheric environment. In this apparatus, hydrogen, methane and ammonia can produce many organic compounds under the action of simulated lightning. About 10% to 15% of the carbon existed in the form of organic compounds, of which 2% was amino acids and the rest includes sugars, lipids and so on. This finding not only broke the traditional idea that inorganic matters in the inanimate system could not be transformed to organic matters, but also aroused the great concern of many scientists about the ultimate problem of the origin of life.

https://doi.org/10.1515/9783110562552-011

11.2 The origin of marine life

11.2.1 The theory of the origin of marine life

The harsh environment of the early Earth (greenhouse effect, frequent volcanic activities, lightning, extraterrestrial meteorites and strong ultraviolet radiation, as well as the lack of atmospheric protection) makes the land environment not suitable for biological survival, while the ocean may be the safest and the most suitable living environment. After archeologists, paleontologists found that, about 3.8 billion years ago, the most primitive life appeared in the ocean, and its structure is very similar to modern bacteria. After approximately 100 million years of evolution, the original cells in the ocean gradually evolved into the original single-cell algae; these original algae experienced hundreds of millions of years of evolution, developed into the original jellyfish, sponge, trilobites, Nautilus and so on. These creatures appeared and eventually experienced the present life on Earth. Meanwhile, water is also an essential component of life activities. Seawater can effectively prevent the damage of ultraviolet rays on life. The famous scholar of origin of life Obanlin found that biological macromolecules dissolved in water can be condensed into small particles, so Obanlin concluded that it is the aggregates that first formed as an entity in the primitive ocean. To sum up, the most primitive life is most likely to be bred in the primitive ocean. The water environment of the original ocean, all kinds of organic compounds accumulating for a long time and other physical and chemical conditions may be the key factors for the emergence of primitive life.

The deep-ocean environment has long been considered a "restricted area" for life, as about 70% of the deep seabed pressure is above 38 MPa, while the Mariana Trench has a pressure of 110Mpa, which is by far the highest pressure found in deep-ocean submarine. It was not until the 1970s that this view was abandoned. In 1977, American scientist Corliss et al. [5] dived into about 2,500 meters deep using Alvin deep submarine in the eastern Pacific Galapagos rift valley; he accidently found the strange spectacular phenomenon of the seabed hydrothermal eruption, which is vividly described as the seabed "black chimney." It contained high concentrations of sulfide surrounding the hydrothermal vents. These surroundings are undoubtedly highly toxic environment for most of the modern creatures; however, this is undoubtedly a vibrant park for creatures. Especially in recent years, more and more new species have been found surrounding the hydrothermal vents and the habit of these creatures and their genic functions are more related to the ancient creatures. Combined with the early Earth environment and the modern seabed hydrothermal environment (high temperature, high pressure, pH gradient changes, etc.), some people think that modern seabed hydrothermal activity is the "living fossil" of studying the Earth's early environment. At the same time, the energy systems such as the special environment of the hydrothermal vent of the seabed, plenty of liquid water, high-intensity hydrostatic pressure, a higher temperature and lack of sunlight make

a great difference to the Earth's surface but here it is more likely to be the birth place of the early life on the Earth. Corliss et al. [6] proposed the scientific hypothesis that "life originated in seabed hydrothermal activity" and made the scientific community excited. Submarine hydrothermal vent and other places have a great difference in physical and chemical environment, such as temperature, pH, pressure, metal ions and small organic molecules and oscillations often occur. Similar to organic solvents, water has a strange physical property in high pressure, that is, it performs strange reaction mechanisms and provides unique physical environment to the origin of life [7,8]. This is especially in submarine volcanic phosphates, such as P_3m, where it can act as a catalyst for life material.

11.2.2 Simulation of chemical evolution of marine life

Since Corliss put forward the scientific hypothesis that "life originated in seabed hydrothermal activity," people have made great discoveries and developments. On the other hand, the hydrothermal regions of different characteristics under the sea are constantly being discovered. In 2000, scientists found a 60 meters high carbonate chimney in the mid-Atlantic ridge [9], located at a very special submarine hydrothermal activity area, with a very wonderful name – lost city hydrothermal area. A large number of microbial communities live around the chimney; the pH of the liquid sprayed from the vent is 9–11, that is alkaline. This provides a natural pH gradient for the interface between the hydrothermal and the seawater (pH 5–6), where the natural proton gradient can be continuously supplied at this alkaline hydrothermal port. At the same time, Martin et al. [10] argued that this form appears earlier than the chemical process of the proton gradient produced by gene control. On the other hand, the study of chemical evolution of simulated hydrothermal environment has also been reported. Foustoukos et al. [11] used CO_2 and H_2O as raw materials to react in a hydrothermal system containing NaCl, FeO/FeO–Cr_2O_3(trace amounts of metal compounds in seabed sediments) (390 °C, 400 bar). The results showed that hydrocarbon compounds were found in two groups of experiments; the amount of hydrocarbons was more in the experimental group containing chromium elements. Bernhardt et al. [12] studied the stability of amino acids at high temperature. The results showed that high temperature (250 °C) accelerated the decomposition of amino acids (glycine and alanine). After 6 hours of high temperature reaction, the amino acids almost completely decomposed. White et al. [13] also demonstrated that high temperature is an important factor in the decomposition of amino acids. The half-life of amino acids in high temperature reactions is less than 7.5 hours. The abovementioned experimental results show that amino acids are unstable under high temperature conditions, but the thermophilic bacteria found in the hydrothermal vent of the seabed can survive at 121 °C, indicating that it must be having the mechanism of anti-heat stability. Amend et al. [14] showed that 20

amino acids were more easily synthesized using CO_2, NH_4^+ and H_2 as raw materials under high temperature conditions (100 °C) than at low temperature conditions (18 °C) by thermodynamic calculations. At the same time in high temperature conditions, 11 kinds of amino acid synthesis reactions were exothermic, whereas in low temperature conditions, all the synthesis reactions are endothermic. Hazen [15] and Franiatte et al. [16] also confirmed this from their side. Hazen added certain amount of minerals gathered around the hydrothermal vent in the reaction solution; by adding these minerals the amino acids can be stabilized at high temperatures (200 °C) for several days. He presumed that the salt ions in the solution were in effect. Franiatte placed adenine at a high temperature of 300 °C; adenine also showed a certain stability.

In addition, Hennet et al. [17] used potassium cyanide, ammonium chloride and formaldehyde as raw materials to catalyze the reaction with minerals (pyrite, pyrrhotite, magnetite and illite) at 10 bar and 150 °C for 24 hours, and got different kinds of amino acids. The results showed that most of the amino acids generated were characterized by d and l types in the experimental group with pyrite, pyrrhotite, magnetite and illite as catalysts; the results showed that most of the amino acids generated are l-form in the experimental group with illite as the catalyst. According to the experimental results, it seems that the importance of minerals as catalysts in the reaction can be seen. Lemke et al. [18] also simulated the reaction of glycine under hydrothermal conditions (160 °C/ 220 °C/260 °C, 20 MPa), resulting in oligopeptide generation.

Amino acids and nucleosides polymerized further into peptides and nucleotides, and this critical step can be catalyzed by polyphosphate or cyclotriphosphate P_3m produced by submarine volcanoes.

The abovementioned studies showed that by simulating high temperature and high pressure conditions of the seabed hydrothermal vent, simple inorganic substances can produce organic matters and simple organic molecules can polymerize into larger organic molecules. At the same time, the result obtained by theoretical calculations also suggested that the high temperature conditions of the seabed hydrothermal vent is more favorable for the formation of organic matter. So, life originated in the ocean is not a groundless hypothesis. The unique ecological environment of submarine hydrothermal area will attract more researchers to enter this area to continue the exploration.

11.3 Phosphate oxygen isotope as a biomarker

Being an important nutrient, phosphorus plays crucial rules for all the lives. Phosphorus takes part in most of the biogeochemical cycles in the form of phosphate. The P–O bond of phosphate is strong (359.8 $kJ.mol^{-1}$), phosphate can only

exchange oxygen isotope with water under the action of organism or enzyme [19, 21]. Therefore, if we know the primary oxygen isotope composition of phosphate ($\delta^{18}O_P$) of the Earth, comparing it with the $\delta^{18}O_P$ value of the samples, we can know whether the phosphorus of the samples had been used by the organism, for example, the existence of lives.

The $\delta^{18}O_P$ is defined as

$$\delta^{18}O_p = \frac{\left(\frac{^{18}O}{^{16}O}\right)_{Sample}}{\left(\frac{^{18}O}{^{16}O}\right)_{SMOW}} - 1 \quad (1)$$

$(^{18}O/^{16}O)_{Sample}$ refers the $^{18}O/^{16}O$ ratio of the PO_4^{3-} in samples. $(^{18}O/^{16}O)$SMOW refers the $^{18}O/^{16}O$ ratio of the Vienna Standard Mean Ocean Water (SMOW); it is equal to 1989.4×10^{-6}.

11.3.1 The oxygen isotope fractionation between water and phosphate

The base of the phosphate oxygen isotope biomarker is that phosphate can only exchange oxygen isotope with water under the action of the organism or enzyme. The dissolved phosphate is the easiest form of phosphate that can exchange oxygen isotope with water. The dissolved phosphate can be transformed into apatite, which is the main form of phosphate present in the rock. Therefore, understanding of the oxygen isotope exchange between water and different forms of phosphate is very important (Figure 11.1).

11.3.1.1 Oxygen isotope fractionation between water and biotic apatite

Biotic apatite is mainly the apatite in the shell, bone or teeth of animals. Based on the $\delta^{18}O_P$ of shells of marine organisms that lived in different water temperature, Longinelli et al. [22] had established the first oxygen isotope fractionation equation of water-phosphate system:

$$T(°C) = 111.4 - 4.3\,(\delta^{18}O_P - \delta^{18}O_W)$$

T refers to temperature; $\delta^{18}O_P$ and $\delta^{18}O_W$ refer to the oxygen isotope composition of phosphate and water, respectively. This equation was proved by studies on the bone [23] and teeth [24] of the fish and the shell of *Lingulid* [25]. For the large quantities and types of samples, and the consensus of results, oxygen isotope fractionation between water and biotic apatite was generally supposed to reach the equilibrium fractionation state, and this equation was widely used.

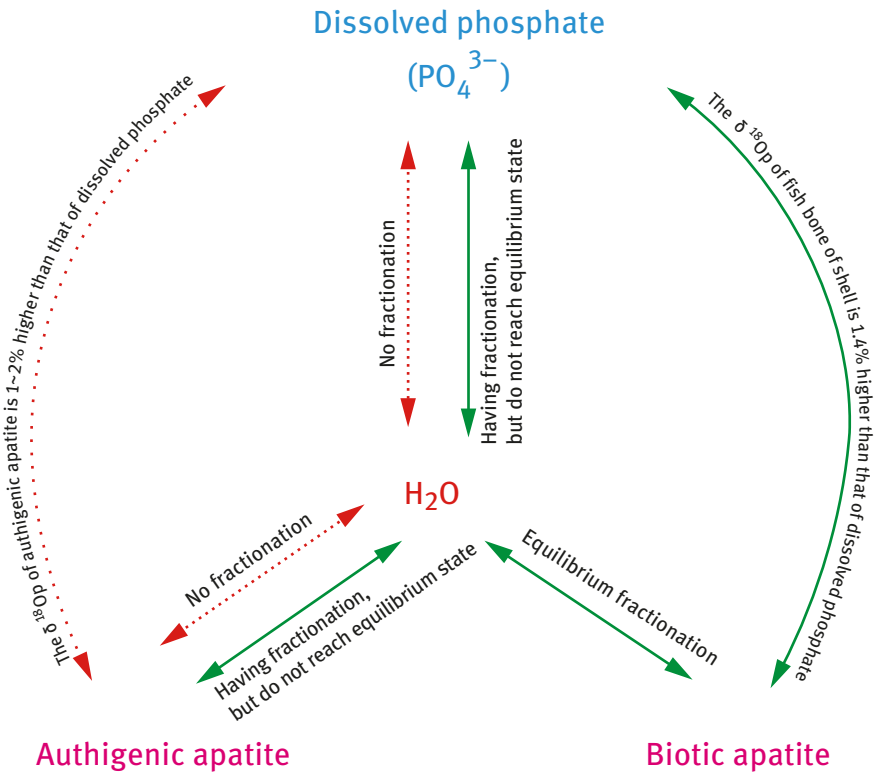


Figure 11.1: The oxygen isotope fractionation of water-phosphate system.
(Dashed lines represent the process that has no activities of the organism or enzyme; solid lines represent the process that has activities of the organism or enzyme)

11.3.1.2 Oxygen isotope fractionation between water and authigenic apatite

Authigenic apatite refers to the apatite that forms in the sediments or during the subsequent diagenetic process. Blake et al. [26] studied the oxygen isotope between water and authigenic apatite by synthesizing apatite in the laboratory. The phosphorus source includes organic phosphorus (RNA) and inorganic phosphorus (KH_2PO_4). And all the samples were separated into two groups: one has microscopic organism (bacteria), while another without any microorganism. The $\delta^{18}O$ of water and formed apatite were measured. For the organic phosphorus source experiments, the results of the group that had organism showed that the organism can quickly decompose the RNA and improve the phosphate concentration. When the $\delta^{18}O_W$ was between −5.5 and −6.6‰, the oxygen isotope fractionation between water and the formed apatite could reach equilibrium fractionation state. While for the other $\delta^{18}O_W$ values, the system deviated from the equilibrium fractionation. They believed 50% of oxygen atoms in the RNA were kept by the formed

phosphate. For the group that had no organisms, the phosphate concentration did not increase (e.g., no RNA was hydrolyzed), and no apatite was formed. For the inorganic phosphorus source experiments, the $\delta^{18}O_P$ of the group that had organisms varied greatly, indicating the organism could lead to oxygen isotope exchange between phosphate and water, but it did not reach the equilibrium fractionation state. While for the group that had no organism, the $\delta^{18}O_P$ was stable, but it was about 2‰ higher than that of the original value; $\delta^{18}O_W$ and temperature had no effects on $\delta^{18}O_P$.

Therefore, although the equation of Longinelli [22] had been proved by many samples of organisms, the results of synthesized apatite have not proved it yet. This may be because of the high phosphate concentration of the study, which led to the incomplete use of the phosphate by the organisms [27].

11.3.1.3 Oxygen isotope fractionation between water and dissolved phosphate

The oxygen isotope of phosphate can trace the metabolism process of phosphate, so the oxygen isotope fractionation between water and dissolved phosphate had been studied for a long time. [28, 29]. These studies showed that without the action of organism or enzyme, dissolved phosphate did not exchange oxygen isotope with water. This conclusion has been proved by many further studies [19, 21].

Organisms take part in almost all the biogeochemical cycling process of phosphorus. Phosphates do not exchange oxygen isotope with water if there is no action of organism or enzyme; this makes the study of the oxygen isotope exchange between dissolved phosphate and water to be simpler and more significant. Blake et al. [27] had found pyrophosphatase (PPase) could speed up the oxygen isotope exchange between water and dissolved phosphate. The system can reach equilibrium state within 500 hours under the temperature range of 5.7°C to 22°C. They had also studied the effects of oxygen isotope of phosphate consumed by *Escherichia coli*. They found that with the increase of biomass, the phosphate concentration decreases from 1050 to 600μM, then it increases and becomes stable at 900μM. During this increase and decrease of phosphate concentration, $\delta^{18}O_P$ showed opposite correlation with the phosphate concentration. When the phosphate concentration was stable at 900μM, the $\delta^{18}O_P$ also became stable. However, $\delta^{18}O_P$ was higher than that of the equilibrium value. They suggested this may be the results of dynamic isotope fractionation of phosphate by the organism.

From these experiments we can see that though phosphate can exchange oxygen isotopes with water under the action of organisms or enzymes, but except PPase, all others had not reached equilibrium fractionation state. This implies that the oxygen isotope exchange between water and dissolved phosphate may depend on the species and the number of organisms and the concentration of phosphate.

11.3.2 The application of oxygen isotope composition of phosphate

11.3.2.1 The primary value of the oxygen isotope composition of phosphate

To use $\delta^{18}O_P$ as a biomarker, we should know the $\delta^{18}O_P$ of the samples whose phosphorus had never been used by an organism (e.g., primary $\delta^{18}O_P$). The meteorite is an ideal candidate. Meanwhile, for the magma coming from deep Earth, $\delta^{18}O_P$ of the igneous rocks can represent the primary $\delta^{18}O_P$ of the Earth too. However, one should note that the mantle can be contaminated by the subduction of crust. At the same time, the upwelling of the magma can also be contaminated by the crust. Therefore, other methods should be used to make sure the samples were not contaminated by the crust materials. There were some $\delta^{18}O_P$ studies of the meteorite and igneous rocks.

Taylor and Epstein [30] had studied the $\delta^{18}O_P$ of the apatite in the San Marcos gabbro and Bonsall Tonalite. The San Marcos gabbro belongs to noritic hornblende gabbro. The apatite constituted 0.29% of it, and its $\delta^{18}O_P$ was 4.1±0.3‰, which was lower than that of other accompanying minerals. The apatite constituted 0.25% of the Bonsall Tonalite, its $\delta^{18}O_P$ was 6.7±0.1‰, which was lower than other accompanying minerals too. Mizota et al. [31] had studied the $\delta^{18}O_P$ of different igneous rocks. They found the $\delta^{18}O_P$ of the four carbonatites was between 0.2 to 5.0‰, and the $\delta^{18}O_P$ of the apatite in two hydrothermal geodes was 2.4‰ and 3.4‰, and the $\delta^{18}O_P$ of the two fresh volcanic ashes was 5.3‰ and 6.2‰.

Greenwood et al. [32] had studied $\delta^{18}O_P$ of merrillite and chlorapatite in Martian meteorite ALH84001 in Los Angeles. Their $\delta^{18}O_P$ ranged from 2.8 to 6.4‰, which was similar to $\delta^{18}O_P$ values of the Earth basalt. The $\delta^{18}O_P$ of the unaltered igneous rocks ranges from 0.2 to 7.0‰. We cannot exclude the possibility of contamination for the samples that have relatively high $\delta^{18}O_P$ values. Therefore, we still do not know the exact primary value of $\delta^{18}O_P$, but the lower the $\delta^{18}O_P$ values are, the lower possibilities of its changes by the organism.

11.3.2.2 The review of the phosphate oxygen isotope biomarker

The use of phosphate oxygen isotope as a biomarker was studied mainly in the laboratory of Professor Blake at Yale. They found the $\delta^{18}O_P$ of dissolved phosphate in underground water had positive correlation with the $\delta^{18}O$ of water; this suggested phosphate had been metabolized by the organism [33]. Based on the $\delta^{18}O_P$ of the metalliferous sediments near the hydrothermal vent of the Red Seamount, the calculated temperatures were 0.8 °C and 60.2 °C, which indicated phosphate was metabolized intensively by the organism [33].

Blake et al. [34] also utilized the isotope of phosphate oxygen to trace the biological activity in the early Earth. The $\delta^{18}O_P$ of the 3.2–3.5-billion-year-old Barberton Greenstone Belt ranged from 9.3‰ to 19.9‰ [34]. The highest values were similar to that of the modern ocean, and were much higher than that of the igneous rocks. This indicated there were strong biological activities in the ocean of that time. More and

more studies had shown there was presence of water in the history of Mars [35], and the soil of Mars also had high phosphorus content [36]. Therefore, the oxygen isotope of phosphate can be an effective biomarker to detect lives on Mars [33].

References

[1] Cronin J R, Pizzarello S, Moore C B. Amino acids in an Antarctic carbonaceous chondrite [J]. *Science*, 1979, 206(4416): 335–337.

[2] Miller S L. Chemical Evolution Molecular Evolution towards the Origin of Living Systems on the Earth and Elsewhere [M]. Calvin, M. Oxford University Press, New York, 1969, 278.

[3] Joyce G. The antiquity of RNA-base devolution [J]. *Nature*, 2002, 418, 214–221.

[4] Miller S L. A production of amino-acids under possible primitive earth conditions [J]. *Science*, 1953, 117: 528–529.

[5] Corliss J B, Dymond J, Gordon L I, Edmond J M et al. Submarine thermal springs on the Galápagos Rift [J]. *Science*, 1979, 203(4385): 1073–1083

[6] Corliss J B, Baross J A and Hoffman S E. An hypothesis concerning the relationship between submarine hot springs and the origin of life on Earth [J]. *Oceanologica Acta*, 1981, 4: 59–69.

[7] Robert M H, Nabil B, Brandes J A et al. High pressure and the origin of life [J]. *Journal of Physics Condensed Matter*, 2002, 14: 11489–11494.

[8] Lee W T, Salji E K H and Bismayer U. Domain wall diffusion and domain wall softening [J]. *Journal of Physics Condensed Matter*, 2003, 15: 1353–1366.

[9] Proskurowski G., Lilley M D, Kelley D S, et al. Low temperature volatile production at the Lost City Hydrothermal Field, evidence from a hydrogen stable isotope geothermometer [J]. *Chemical Geology*, 2006, 229(4): 331–343.

[10] Martin, W. and M.J. Russell. On the origin of biochemistry at an alkaline hydrothermal vent [J]. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 2007, 362: 1887–1926.

[11] Foustoukos D I and Seyfried W E. Hydrocarbons in hydrothermal vent fluids: the role of chromium-bearing catalysts [J]. *Science*, 2004, 304(5673): 1002–1005.

[12] Bernhardt G., Ludemann H D, Jaenicke R et al. Biomolecules are unstable under “black smoker” conditions [J]. *Naturwissenschaften*, 1984, 71(11): 583–586.

[13] White, R.H., Hydrolytic stability of biomolecules at high temperatures and its implication for life at 250 °C [J]. *Nature*, 1984, 310(5976): 430–432.

[14] Amend J P and Shock E L. Energetics of amino acid synthesis in hydrothermal ecosystems [J]. *Science*, 1998, 281(5383): 1659–1662.

[15] Hazen R M. Life’s rocky start [J]. *Scientific American*, 2001, 284(4): 76–85.

[16] Franiatte M, Richard L, Elie M, et al. Hydrothermal stability of adenine under controlled fugacities of N_2, CO_2 and H_2 [J]. *Origins of Life and Evolution of Biospheres*, 2008, 38(2): 139–148.

[17] Hennet R C, Holm N and Engel M. A biotic synthesis of amino acids under hydrothermal conditions and the origin of life: a perpetual phenomenon [J]? *Naturwissenschaften*, 1992, 79(8): 361–365.

[18] Lemke K H, Rosenbauer R J and Bird D K. Peptide synthesis in early earth hydrothermal systems [J]. *Astrobiology*, 2009, 9(2): 141–146.

[19] Lecuyer C, Grandjean P and Sheppard S M F. Oxygen isotope exchange between dissolved phosphate and water at temperatures≤135°C: Inorganic versus biological fractionations [J]. *Geochimica et Cosmochimica Acta*, 1999, 63(6): 855–862.

[20] O'Neil J R, Vennemann T W and McKenzie W F. Effects of speciation on equilibrium fractionations and rates of oxygen isotope exchange between (PO_4) aq and H_2O [J]. *Geochimica et Cosmochimica Acta*, 2003, 67(17): 3135–3144.
[21] Liang Y H and Blake R E. Oxygen isotope fractionation between apatite and aqueous-phase phosphate: 20–45°C [J]. *Chemical Geology,* 2007, 238(1–2): 121–133.
[22] Longinelli A, Nuti S. Revised phosphate-water isotopic temperature scale [J]. *Earth and Planetary Science Letters*, 1973, 19(3): 373–376.
[23] Longinelli A, Nuti S. Oxygen isotope measurements of phosphate from fish teeth and bones [J]. *Earth and Planetary Science Letters,* 1973, 20(3): 337–340.
[24] Kolodny Y, Luz B, Navon O. Oxygen isotope variations in phosphate of biogenic apatites I. Fish bone apatite – rechecking the rules of the game [J]. *Earth and Planetary Science Letters*, 1983, 64(3): 398–404.
[25] Lecuyer C, Grandjean P and Emig C C. Determination of oxygen isotope fractionation between water and phosphate from living lingulids: Potential application to paleoenvironmental studies [J]. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 1996, 126(1–2): 101–108.
[26] Blake R E, O'Neil J R and Garcia G A. Oxygen isotope systematics of microbially mediated reactions of phosphate: I. Degradation of organophosphorus compounds [J]. *Geochim et Cosmochim Acta*, 1997, 61(20): 4411–4422.
[27] Blake R E, O'Neil J R and Surkov A. Biogeochemical cycling of phosphorus: insights from oxygen isotope effects of phosphoenzymes [J]. *American Journal of Science*, 2005, 305(6–8): 596–620.
[28] Blumenthal E and Herbert J B M. Interchange reactions of oxygen. 1. Interchange of oxygen between water and potassium phosphate in solution [J]. *Transactions of the Faraday Society*, 1937, 33: 849–852.
[29] Winter E R S, Carlton M and Briscoe H V A. The interchange of heavy oxygen between water and inorganic oxy-anions [J]. *Journal of the Chemical Society (Resumed)*, 1940, 131–138.
[30] Taylor H P and Epstein S. Relationship Between O18/O16 Ratios in Coexisting Minerals of Igneous and Metamorphic Rocks Part 1: Principles and Experimental Results [J]. *Geological Society of America Bulletin*, 1962, 73(4): 461–480.
[31] Mizota C, Domon Y and Yoshida N. Oxygen isotope composition of natural phosphates from volcanic ash soils of the Great Rift Valley of Africa and east Java, Indonesia [J]. *Geoderma*, 1992, 53(1): 111–123.
[32] Greenwood J P, Blake R E and Coath CD. Ion microprobe measurements of 18 O/16 O ratios of phosphate minerals in the Martian meteorites ALH84001 and Los Angeles [J]. *Geochimica et Cosmochimica Acta,* 2003, 67(12): 2289–2003.
[33] Blake R E, Alt J C, Martini A M. Oxygen isotope ratios of PO4: an inorganic indicator of enzymatic activity and P metabolism and a new biomarker in the search for life [J]. *Proceedings of the National Academy of Sciences*, 2001, 98(5): 2148–2153.
[34] Blake R E, Chang S J and Lepland A. Phosphate oxygen isotopic evidence for a temperate and biologically active Archaean ocean [J]. *Nature*, 2010, 64(7291): 1029–1032.
[35] Ojha L, Wilhelm M B, Murchie S L, McEwen A S et al. Spectral evidence for hydrated salts in recurring slope lineae on Mars [J]. *Nature Geoscience*, 2015, 8(11): 829–832.
[36] Adcock C T, Hausrath E M and Forster P M. Readily available phosphate from minerals in early aqueous environments on Mars [J]. *Nature Geoscience*, 2013, 6(10): 824–827.