 doi:10.20944/preprints202206.0370.v1

Review

How Did Life Emerge in Chemically Complex Messy Environments?

Kenji Ikehara[1,2,3,*]

[1] G&L Kyosei Institute (in The Keihanna Academy of Science and Culture (KASC)), Keihanna interaction plaza, Lab. Wing 3F, 1-7 Hikaridai, Seika-cho, Souraku, Kyoto 619-0237, Japan.

[2] International Institute for Advanced Studies, Kizugawadai 9-3, Kizugawa, Kyoto 619-0225, Japan.

[3] Professor Emeritus, Nara Women's University, Japan

* Correspondence: ikehara@cc.nara-wu.ac.jp; Tel.: +81-774-73-4478

Abstract:

One of the problems, which make it difficult to solve the mystery of the origin of life, would be how life emerged in chemically complex messy environments on the primitive Earth. It is considered that three main points contributed to open the way to the emergence of life. (1) A characteristic inherent in [GADV]-amino acids, which are easily produced with prebiotic means. (2) Protein 0^{th}-order structure or [GADV]-amino acid composition generating water-soluble globular protein with some flexibility, which can be produced even by random joining of [GADV]-amino acids. (3) Formation of versatile [GADV]-microspheres, which can grow, divide, proliferate even without genetic system, was the emergence of proto-life. (4) [GADV]-microspheres with a higher proliferation ability than others could be selected. The proto-Dawin evolution made it possible to proceed forward to creation of the core life system composed of $(GNC)_n$ gene, anticodon stem-loop tRNA or AntiC-SL tRNA (GNC genetic code) and [GADV]-protein. (5) Eventually, the first genuine life with the core life system emerged. Thus, the formation processes of [GADV]-protein and $(GNC)_n$ gene in chemically complex messy environments were the steps to the emergence of genuine life.

Keywords: GADV hypothesis; origin of life; protein 0^{th}-order structure; origin of protein; [GADV]-microsphere; origin of gene: the core life system

1. Introduction

Human beings have tried to know for many years how and from where we have come. In other words, human beings have interested in the origin of life. In addition, there is another significance in the life-origin research. For example, answers to the fundamental questions about gene and protein, "why is the kind of nucleotides or bases of RNA and DNA four?" and similarly "why is the kind of amino acids of modern proteins 20?", could

 doi:10.20944/preprints202206.0370.v1

be obtained, if the mystery of the origin of life were solved. However, irrespective of strenuous efforts of many researchers, the mystery of the origin of life remains unsolved still now. The main reasons would be as follows.

A. Difficulties to elucidate the establishment process of the fundamental life system

The most important point for solving the mystery of the origin of life would be to make clear the establishment process of the fundamental life system composed of six members (gene, genetic code, tRNA, metabolism, cell structure and protein) [1; Chapter 2] However, it has been difficult to understand the establishment process because of the following reasons.

1. The "chicken-egg relationship" between gene and protein: The so-called "chicken-egg relationship" has made it difficult to solve the mystery of the origin of life for many years. In such situation, the RNA world hypothesis, assuming that the first life arose from RNA world formed by self-replication of RNA, was proposed by Gilbert in 1986 [2]. However, it would be principally impossible to solve the mystery from the standpoint of the hypothesis, because any gene encoding a mature protein never be formed in the absence of a target protein or in the absence of protein, even if RNA could be produced on the primitive Earth.

On the other hand, I have proposed the [GADV]-protein world hypothesis, in short GADV hypothesis, about 20 years ago [1,3,4], assuming that the first life emerged from [GADV]-protein world, which was formed by pseudo-replication of [GADV]-proteins [5]. I believe now that the mystery could be solved based on the GADV hypothesis. [GADV] means four amino acids, Gly [G], Ala [A], Asp [D] and Val [V]. Square bracket, [ ], is used for discrimination of one letter symbol of amino acids, especially Gly [G], Ala [A] from one letter symbol of nucleobases, guanine G, adenine A.

2. The emergence of life under random processes: Any affairs happened on the primitive Earth should proceed under random processes. However, it is easily supposed that gene and protein with an ordered sequence never be formed through a random process at one stroke, because sequence diversities of gene encoding a protein composed of 100 amino acids and the protein itself are extraordinary large as $(4^3)^{100} = \sim 10^{180}$ and $20^{100} = \sim 10^{130}$, respectively [1; Chapter 3, 6]. On this problem, I believe that the question, how a gene could acquired genetic information of a mature protein, could be solved, if the question is considered from the standpoint of the GADV hypothesis [1; Chapter 8], as explained later in detail.

B. The problem of research strategy based on experiments:

1. The mystery of the origin of life could not be solved only with experiments: Needless to say, experiments are quite important or crucial, even in studying on the origin of life. However, it would be impossible to solve the mystery only with experiments, because

 doi:10.20944/preprints202206.0370.v1

affairs happened on the primitive Earth about 4 billion years ago never be reproduced by experiments carried out in a present laboratory.

2. Establishment process of the core life system composed of gene, genetic code (tRNA) and protein [1; Chapter 2], could not be solved only with bottom-up approaches: It is undoubtedly important in order to make clear the origin of life to answer to the questions, what affairs, which led to the emergence of life, happened on the primitive Earth. Therefore, many researchers have carried out the studies according to bottom-up approaches to know for many years, where and what kind of organic compounds were produced on the primitive Earth. However, it would be probably impossible to solve the mystery of the origin of life, because the establishment process of the core life system would never be made clear only with the bottom-up approaches, unless newly-born life or RNA/DNA having the most primitive genetic information for protein synthesis can be found from old rocks about 4 billion years ago etc.

3. Difference between an experimental condition in a laboratory and the situation on the primitive Earth: Experiments for confirming nucleotide synthesis on the primitive Earth have been carried out by many researchers and the results showing that nucleosides were actually produced with ribose and nucleobases are frequently reported in scientific journals [7-12]. However, it seems to me that it would be impossible to apply the results to the real affairs happened on the primitive Earth, because the experimental conditions, such as concentration and purity of reactants, are always far from the situation on the primitive Earth. Therefore, I consider that a parallel use of both the bottom-up approaches, which carried out thus far, and top-down approaches, for example, using data-base analyses of modern genes and proteins, should hold one of the keys for solving the origin of life [13].

C. The problem about how gene and protein, which are composed of the respective small kinds of components, could be produced in chemically complex messy environments on the primitive Earth

As described in "Special Issue Information" of the special issue (Origin of Life in Chemically Complex Messy Environments), as "*Considering the prebiotic Earth four billion years ago (a messy atmosphere, in other words), a chaotic mélange of diverse starting materials appears realistic*", there is another problem, which makes it difficult to solve the mystery of the origin of life. Biopolymers as gene and protein, which are composed of only four kinds of nucleotides and twenty kinds of amino acids, respectively, are used in extant organisms. However, such biopolymers must be formed in chemically complex messy environments on the primitive Earth. Therefore, it should become a problem how such biopolymers using a small kinds of components could be formed in the messy envionments on the primitive Earth. When the mechanism producing gene and protein, which use the respective small number of components, was acquired by something, the something became the first life. Therefore, understanding the steps to the emergence of life should lead to discovery of a correct answer to the question how gene and protein could be produced in chemically complex messy environments.

 doi:10.20944/preprints202206.0370.v1

In this paper, I would like to discuss the processes, how biopolymers composed of rather a small kinds of the respective simple monomers were formed in the chemically messy environments, from the standpoint of GADV hypothesis [1,3,4] . For the purpose, it is necessary to understand through what processes the first life emerged on the primitive Earth. Then, I first explain only about the main points below, how the GADV hypothesis could be proposed, as already described about my idea in detail on the origin of life in the book "Towards revealing the origin of life -Presenting the GADV hypothesis", which was published last year from Springer [1]. Thereafter, it is explained how gene and protein using a small kinds of the respective components were formed during repeated random reactions in chemically messy environments, and I would like to give an answer to the third question (C), how genes and proteins could be produced using a small kinds of the respectively selected monomers from the messy environments.

2. The key for solving the mystery of the origin of life (Protein 0^{th}-order structure: [GADV]-amino acids)

The most important concept in the GADV hypothesis is one of the protein 0^{th}-order structures or [GADV]-amino acid composition [14]. Then, I first explain the significance of protein 0^{th}-order structure in solving the origin of life (Figure 1).

Consider here the problem based on the GADV hypothesis, how a gene, which encodes one amino acid sequence of a mature protein, was created. It is well known that it is impossible to produce one mature protein like a precise polymer machine through random joining of even of simple four [GADV]-amino acids at one stroke, because the amino acid sequence diversity of a protein composed of only four kinds of one hundred [GADV]-amino acids is extraordinary large (4^{100} = ~10^{60}) [1: Chapter 3].

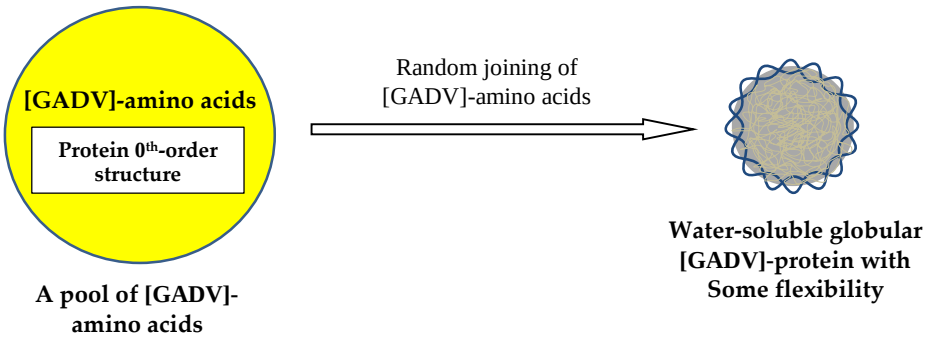


Figure 1. Polypeptide chain, which is obtained by joining of amino acids randomly selected out from a pool (a protein 0^{th}-order structure) containing roughly equal amounts of [GADV]-amino acids,

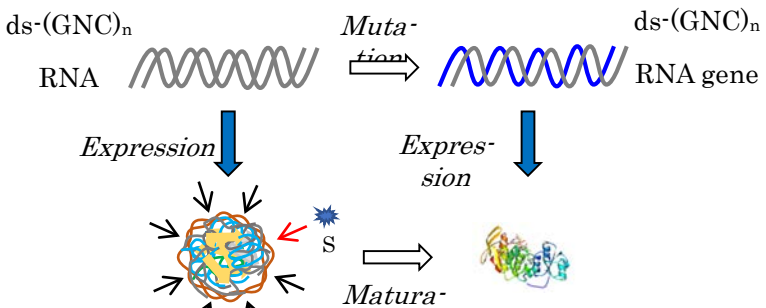
 doi:10.20944/preprints202206.0370.v1

should be folded into a water-soluble globular structure with some flexibility. The protein, actually aggregates of [GADV]-peptides, has a pluripotency that makes it possible to exhibit many catalytic activities owing to the structure flexibility. Wavy lines surrounding a gray circle and thin yellow curved lines within the circle mean flexible structure of an immature [GADV]-protein.

How were genes encoding a mature protein formed? For the purpose, one immature water-soluble globular protein with some flexibility, which is produced by expression of one of double-stranded (ds)-RNA encoding one essentially random [GADV]-amino acid sequence in the protein 0^{th}-order structure, is indispensable for formation of the gene (Figure 2). Namely, every gene encoding a mature protein has been formed as a result of maturation from an immature or incomplete water-soluble globular protein with some flexibility, which generates various weak catalytic activities or demonstrates pluripotency [1; Chapter 3], to a mature protein having a rigid structure and a high catalytic activity.

This means that formation of a mature protein always require an immature protein, because the formation of a mature protein must be always led by elevation of a weak catalytic activity of the immature protein as easily understood if the relationship between a key and a key hole is considered (Figure 2). This is one of the reasons why the mystery of the origin of life has not been solved by RNA world hypothesis, because any RNA never acquire genetic information for a mature protein synthesis independently of an immature protein, even if RNA could be first produced by a random process or random joining of nucleotides in prebiotic environments.

As described above, both an immature [GADV]-protein and a (ds)-RNA are required to form the first RNA gene (Figure 2). At the present, an essentially random codon sequence similar to random $(SNS)_n$ codon sequence or a non-stop frame on antisense strand of a GC-rich gene is used as the field for creating an entirely new gene [15]. On the primitive Earth, immature [GADV]-proteins were produced using a [GADV]-amino acid sequence encoded by a random $(GNC)_n$ sequence on one of two RNA strands (Figure 2). The reason, why it was possible, is because one amino acid sequence randomly selected from [GADV]-amino acid pool (protein 0^{th}-order structure) is essentially the same as one amino acid sequence, which was arranged by a random $(GNC)_n$ RNA sequence. Thus, the GADV hypothesis, which I have proposed, is an idea based on one of the protein 0^{th}-order structures or [GADV]-amino acid composition.

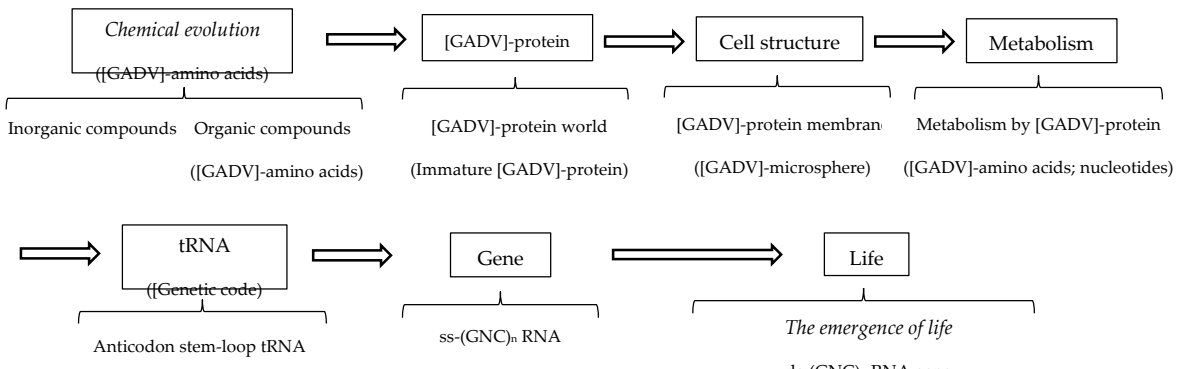


Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 June 2022 doi:10.20944/preprints202206.0370.v1

Figure 2. Every mature protein must be generated by maturation of an immature protein, which was produced by expression of an RNA strand encoding random $(GNC)_n$ sequence for the immature protein. The reason is because a substrate binding site (a key hole) must be formed as accumulating appropriate mutations to adjust the site to fit closely to a substrate (a key). For the purpose, protein 0^{th}-order structure is indispensable to produce an immature protein with some flexibility. Brown wavy lines surrounding an immature protein and blue curved lines mean flexible structure of the immature [GADV]-protein. Curved gray lines and a blue line indicate random $(GNC)_n$ RNA strands and a $(GNC)_n$ RNA gene encoding a mature [GADV]-protein, respectively.

3. Possible steps from chemical evolution to the emergence of the first life

Next, consider the steps concretely to the emergence of the first life on the primitive Earth according to GADV hypothesis [1] (Figure 3). The reason, why the steps to the emergence of life must be described here, is because the steps are intimately related to the establishment process of the core life system composed of gene, genetic code (tRNA) and protein and to the formation process of protein composed of four kinds of [GADV]-amino acids, which were encoded by $(GNC)_n$ RNA gene composed of four nucleotides. In other words, the emergence of a genuine life was intimately related to the formation process of $(GNC)_n$ RNA gene and [GADV]-protein using the respective small number of components in chemically messy environments on the primitive Earth.



 doi:10.20944/preprints202206.0370.v1

Figure 3. Possible steps from chemical evolution to the emergence of life, which is deduced from [GADV]-protein world hypothesis (GADV hypothesis). All of the steps are related to [GADV]-amino acids or [GADV]-protein, including GNC genetic code and $(GNC)_n$ gene both of which encode [GADV]-amino acids and [GADV]-protein, respectively. The steps to the emergence of life can be reasonably explained with the GADV hypothesis assuming that life emerged as piling up the six members (protein, cell structure, metabolism, tRNA, genetic code and gene) in order of from [GADV]-protein to $(GNC)_n$ gene encoding a mature [GADV]-protein.

Then, I explain the main object of this article about the way how biopolymers (gene and protein) composed of a small number of the respective monomers were formed in chemically messy environments on the primitive Earth.

In this Section, the steps from chemical evolution to the emergence of the first genuine life, which was equipped with the fundamental life system, are discussed as dividing them into three (Table 1): The first part: from chemical evolution to formation of [GADV]-microsphere (proto-life). The second part: from formation of the [GADV]-microsphere to formation of AntiC-SL RNA. The third part: from AntiC-SL RNA to the emergence of the first genuine life. The reason, why it is discussed as dividing the processes into three parts, is because the aspect of selection or of evolution changed before and after formation of [GADV]-microsphere and also the reason, why the second and third parts are separated, is because the use of nucleotides is restricted for the first time into four, A, G, U and C, by formation of AntiC-SL RNA, which is folded with base pairs, A-U and G-C. On the other hand, the number of amino acids could not be still restricted into four at the time point. The number of amino acids could not be restricted until the establishment of the core system composed of gene, tRNA (genetic code) and protein.

Table 1: Steps from chemical evolution to the emergence of life, which are discussed in this article as dividing into three parts, I, II and III.

 doi:10.20944/preprints202206.0370.v1

Part I: From chemical evolution to formation of [GADV]-microsphere

1. [GADV]-amino acid synthesis with prebiotic means
2. Formation of [GADV]-microspheres ([GADV]-protein world)
3. Selection and evolution of [GADV]-microsphere

****

Part II: After formation of [GADV]-microsphere to formation of AntiC-SL tRNAs

1. Formation of proto-metabolism in [GADV]-microsphere

   [GADV]-amino acid and [GADV]-peptide syntheses, Nucleotide synthesis with immature [GADV]-proteins (actually aggregates of [GADV]-peptides)
2. Formation of AntiC-SL tRNAs

In addition, only random processes should be naturally repeated before the first genuine life emerged or in the absence of any gene encoding an ordered amino acid sequence of a [GADV]-protein. Therefore, it is also essential to make clear how [GADV]-proteins with an ordered amino acid sequence were formed or how genes encoding the amino acid sequence were formed during repeating random processes in order to solve the mystery of the origin of life.

Part I. Steps from Chemical Evolution to Formation of [GADV]-microsphere ([GADV]-protein world)

Many studies on chemical evolution have carried out to make clear what kinds of organic compounds can be synthesized from inorganic compounds under what conditions [16, 17]. In the studies, it was investigated mainly about what kinds of biomolecules, such as amino acids, sugars, nucleobases etc., which are necessary for the first life to emerge, can be synthesized.

As easily understood from the previous studies, messy organic compounds including various organic acids and amines in addition to various non-natural amino acids should be produced with prebiotic means. Steps to the emergence of life proceeded through reaction processes with the compounds selected from messy organic compounds. However, it would be quite difficult to select out only necessary organic compounds, for

 doi:10.20944/preprints202206.0370.v1

example only [GADV]-amino acids from messy mixtures of organic compounds, as advocated in the GADV hypothesis (Figure 4). In this Part I, I will explain a principle, which can advance toward a solution to the difficult problem from the viewpoint of GADV hypothesis.

1. Preferential synthesis of [GADV]-amino acids with prebiotic means

In the previous studies, it is confirmed that simple organic compounds as amino acids and nucleobases with a small number of carbon atoms were synthesized with various prebiotic means as described below.

(1) Electric discharge into primitive atmosphere (Miller's type experiments) [16, 17]

(2) Catalytic reactions on pyrite, clay, hydrothermal vents and so on [18-20].

(3) Organic synthesis by heavy bombardments of meteorites or asteroids [21, 22].

(4) Introduction of organic compounds by meteorite, asteroids, space dusts and so on from space [23, 24]

Thus, various organic compounds including especially oxygen atom(s) were synthesized through physical and physico-chemical reactions and messy compounds were introduced from space and accumulated on the primitive Earth.

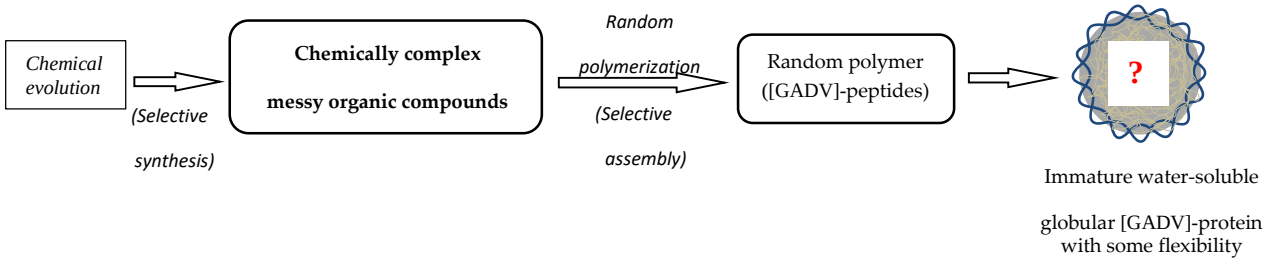


Figure 4. Organic compounds, which were synthesized from inorganic compounds on the primitive Earth with various prebiotic means, were, naturally, chemically complex messy organic compounds. Therefore, a pool containing only [GADV]-amino acids did not exist on the primitive Earth as expected by the GADV hypothesis. Therefore, one important question, how immature water-soluble globular [GADV]-proteins, which is a prerequisite in the GADV hypothesis, could be produced, arises. Presenting an answer to the question how [GADV]-amino acids were selected and immature [GADV]-protein could be formed, is a main purpose of this article.

 doi:10.20944/preprints202206.0370.v1

Amino acids were produced with organic compounds, which were selectively synthesized and accumulated at a large amount on the primitive Earth through prebiotic means. Note that, at this time point, some kind of selection among organic compounds had been carried out during synthetic processes of the organic compounds based on nature of chemical compounds themselves used as reactants, as described below.

1. Preferential synthesis of α-amino acids

Amino acids were produced with prebiotic methods more easily than fatty acids and hydrocarbons, of which constituent atoms are connected chiefly with carbon-carbon and carbon-hydrogen bonds. In fact, it is described by Miller in his book that there is no good method for fatty acid synthesis with prebiotic means [16]. Furthermore, it is well known that α-amino acids were obtained at larger amounts than β-amino acids in Miller's experiments [16]. These results indicate that α-amino acids were rather easily synthesized with prebiotic means and accumulated at large amounts.

2. Direct random joining of [GADV]-amino acids

Next, the mechanism or principle is described about how only [GADV]-amino acids could be eventually selected out at relatively high rate for primeval protein synthesis in the messy mixture of various organic compounds containing non-natural amino acids. The reasons why α-amino acids could be preferentially selected for synthesis of protein are as follows.

[GADV]-proteins, actually aggregates of [GADV]-peptides, could be also produced with prebiotic means in messy organic compounds, which accumulated at large amounts on the primitive Earth, although the proteins were incomplete in the sense that various amino acids other than [GADV]-amino acids were contained in the proteins. On the contrary, nucleotides, which are necessary to produce RNA, could not be synthesized at large amounts with prebiotic means, because of the complex chemical structures of nucleotides. It is known in the previous studies that nucleotides could not be produced by Miller's experiments and nucleotides have not been found in meteorites [16, 17].

1. Preferential polymerization of [GADV]-amino acids

Amino acids should be selectively linked with each other during repeated wet-drying processes, because amino acids have both positive and negative charges in the molecules to easily associate in water and to make peptide bond between two amino acids. Nevertheless, various organic compounds other than [GADV]-amino acids were naturally incorporated into [GADV]-proteins or [GADV]-peptide aggregates during direct random joining of [GADV]-amino acids.

2. Preferential association of [GADV]-peptides containing Val

All four [GADV]-amino acids having rather simple structure were easily synthesized with prebiotic means. However, the synthetic amount of Val, which has the more complex

 doi:10.20944/preprints202206.0370.v1

molecular structure, should be much less than other three amino acids, Gly, Ala, and Asp. Nevertheless, the less amount of Val could be compensated through formation of [GADV]-peptide aggregates owing to a large hydrophobicity of Val, because peptides containing a more amount of Val could be preferentially associated with each other through the large hydrophobic interactions. This also contributed to formation of more active immature [GADV]-proteins.

3. Preferential synthesis of [GADV]-amino acids by immature [GADV]-proteins

It can be considered that various organic compounds, especially [GADV]-amino acids, could be preferentially synthesized using simple organic compounds with functional groups as glyoxylate and pyruvate, even by immature [GADV]-proteins with weak catalytic activities, which were produced by random joining of [GADV]-amino acids under the protein 0^{th}-order structure [1; Chapter3, 5]. Contrary to that, it must be difficult to synthesize hydrocarbons with the immature [GADV]-proteins. It is considered that the selective synthesis of [GADV]-amino acids by immature [GADV]-proteins, which were performed before cell structure formation, advanced further the steps to the emergence of life at a higher rate than before.

On the other hand, many inactive and useless [GADV]-peptides should be also produced during direct random joining of [GADV]-amino acids, because of incorporation of various organic compounds into the peptides. In the reaction, [GADV]-peptides with a sufficiently high catalytic activity could be always produced as a result of a wide distribution of [GADV]-peptides, which were synthesized through random process, although the formation rate of active [GADV]-proteins might be low [25]. This made it possible to proceed towards the emergence of life at a particularly faster rate than the era of chemical evolution without immature [GADV]-peptide catalysts.

4. Possible exclusion of non-natural amino acids owing to protein 0^{th}-order structure

Various α-amino acids and β-amino acids, other than [GADV]-amino acids, such as 2-aminobutylic acid (2-ABA), 2-aminopentanoic acid, β-alanine and so on, should be also produced with prebiotic means. Therefore, various amino acids other than [GADV]-amino acids would be also incorporated into immature [GADV]-proteins during polymerization of [GADV]-amino acids, because non-natural amino acids could not be effectively excluded during the simple polymerization among amino acids with both positive and negative charges in the molecule. Thus, [GADV]-microspheres were formed through such processes.

Part II. Steps from Formation of [GADV]-microsphere to AntiC-SL RNA formation

1. Significance of cell structure as [GADV]-microsphere for facilitating chemical evolution

1. Incorporation of various organic compounds into [GADV]-microsphere

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 June 2022 doi:10.20944/preprints202206.0370.v1

After sufficient amounts of [GADV]-amino acids accumulated on the primitive Earth, [GADV]-microspheres were formed for example by repeated wet-drying processes in depressions of rocks on the primitive Earth [1]. The [GADV]-microspheres inevitably contained a large amounts of [GADV]-peptides in the cell structure so that [GADV]-peptides were synthesized by immature [GADV]-proteins in the microspheres. The supposition that the immature [GADV]-proteins could synthesize [GADV]-peptides is supported by the facts that even Gly-Gly and Gly-Gly-Gly have peptide synthetic activity [26]. Formation of [GADV]-microsphere made it possible to hold oligomeric [GADV]-peptides, which were synthesized in the microsphere, owing to semipermiable [GADV]-protein membrane. Accumulation of [GADV]-petides in the [GADV]-microsphere generated a higher osmotic pressure to induce further incorporation of low molecular weight organic compounds as glyoxylate and pyruvate into the microsphere (Figure 5).

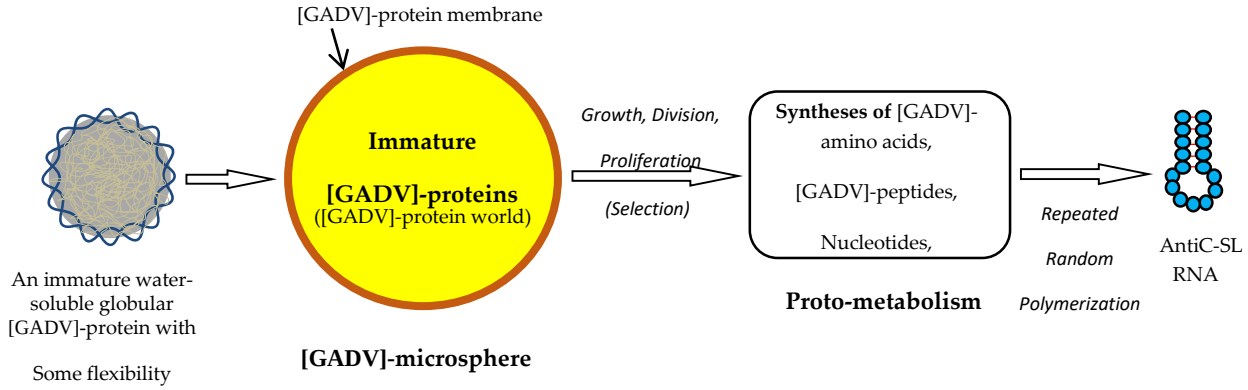


Figure 5. Possible steps from formation of immature [GADV]-proteins to formation AntiC-SL RNA [1]. [GADV]-microspheres surrounded by [GADV]-protein membrane could be formed with [GADV]-peptides during repeated wet-drying processes. [GADV]-protein world was formed in the [GADV]-microsphere. [GADV]-amino acids and nucleotides were synthesized by immature but pluripotent [GADV]-protein catalysts through proto-metabolism in the microsphere. Successively, AntiC-SL RNA was produced by repeated random polymerization of nucleotides and its degradation. The core life system was established in [GADV]-microspheres with a higher proliferation ability were generated through the processes and, eventually, the first life arose on the primitive Earth.

 doi:10.20944/preprints202206.0370.v1

However, such proteins, into which amino acids other than [GADV]-amino acids were incorporated with a higher rate, should be excluded gradually, because [GADV]-microspheres using [GADV]-peptides composed of a higher rate of [GADV]-amino acids should be selected at a high probability during proliferation followed by evolution. The reason, why the selection became possible, is because [GADV]-amino acid composition is one of protein 0^{th}-order structures and incorporation of non-natural amino acids into [GADV]-proteins caused malfunction of the [GADV]-proteins. For the same reason, 2-ABA was excluded from natural amino acids as [GADV]-amino acids in order to avoid duplicate use of Ala and 2-ABA, both of which are α-helix-forming amino acid [27]. The most important aspect of [GADV]-microsphere formation would be the growth and proliferation, which was induced by a high osmotic pressure, as shown (Figure 6).

2. Selection of [GADV]-microspheres with a high proliferation ability

Actually, at first incomplete [GADV]-peptides containing non-natural amino acids would be produced at a high probability. However, [GADV]-microspheres containing non-natural amino acids at a lower rate, could grow, proliferate and evolve faster than others to leave more descendants even before establishment of the genetic system, because [GADV]-microspheres containing at higher rate of [GADV]-amino acids could acquire [GADV]-proteins with a higher function necessary to proceeding to the emergence of life (Figure 6). Thus, [GADV]-microspheres containing at less amounts of non-natural amino acids than others were consequently selected and proliferated. Furthermore, formation of [GADV]-microspheres or [GADV]-protein membrane made it possible to protect dissipation of [GADV]-amino acids into environments, because immature [GADV]-proteins, which was produced with [GADV]-amino acids, could not oozed out through the [GADV]-protein membrane. This also contributed to a more efficient chemical evolution.

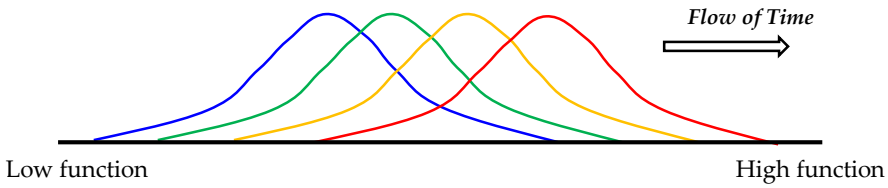


Figure 6. Evolution of [GADV]-microspheres without the genetic system. Random joining of amino acids carried out in protein 0^{th}-order structure inevitably generates [GADV]-peptides or [GADV]-proteins with a large distribution because of the random process. Therefore, at least a part of [GADV]-proteins had always sufficiently high catalytic activities, which should be effective to proceed to the emergence of life. Thus, [GADV]-proteins having a higher catalytic activity than before could be generated in the microspheres. The steps to the emergence of life were the process for acquirement of [GADV]-proteins with a higher catalytic activity than before, step-by-step.

 doi:10.20944/preprints202206.0370.v1

3. Growth, division, proliferation and death of [GADV]-microspheres

[GADV]-microsphere with a higher ability for growth, division and proliferation could be consequently selected out among many [GADV]-microspheres as described above. The selected [GADV]-microspheres could leave more descendants and evolved further, even if the microspheres did not hold any genetic system. Contrast to that, many other microspheres, which were defeated in the struggle for existence, died and left many dead bodies for example in depressions of rocks on the primitive Earth. However, those dead bodies were reused for growth and prosperity of the selected [GADV]-microspheres. This situation is similar to the fact observed on the present Earth that components of withered plants and dead bodies of animals are usually reused by presently living organisms, after those were degraded by various organisms including bacteria.

4. Formation of proto-metabolic pathways for [GADV]-amino acid synthesis

As described above, [GADV]-amino acids having a sufficiently high stability were easily produced and accumulated at large amounts on the primitive Earth. In addition, it was easy to refill [GADV]-amino acids, even if the amino acids were exhausted upon consumption for growth, because [GADV]-amino acids could be synthesized with simple organic compounds, as glyoxylate and pyruvate, which accumulated on the primitive Earth at large amounts and could be easily supplied from the environments into the microspheres (Figure 5). Thus, [GADV]-amino acids were optimal compounds for advancing the steps to the emergence of life on the primitive Earth.

Even such [GADV]-amino acids, which supported proliferation of [GADV]-microspheres, would be depleted from environments and proliferation of the [GADV]-microspheres should terminate soon after the deprivation, because of the exponential growth of [GADV]-microsperes. Only one way to avoid the situation was construction of the proto-metabolism for synthesis of [GADV]-amino acids in the microspheres growing exponentially [1; Chapter 5].

Proto-metabolic reactions using immature [GADV]-proteins started in [GADV]-microspheres to produce [GADV]-amino acids and [GADV]-peptides, just after formation of [GADV]-microspheres. It would be supported by the facts that [GADV]-amino acids can be synthesized using glyoxylate and pyruvate as starting materials in a few reaction steps of modern metabolic pathways (Figure 7 (A)). Inversely, the cycle of growth, division and proliferation of [GADV]-microspheres would terminate, if the proto-metabolic pathways were not formed in [GADV]-microspheres and the sufficient osmotic pressure could not be maintained. Therefore, it can be considered that only [GADV]-microspheres, which had a high synthetic ability of [GADV]-peptides, were evolutionally selected and could leave more descendants than others. From the considerations, it can be concluded that

 doi:10.20944/preprints202206.0370.v1

formation of cell structures, which were indispensable to selection and evolution and in which [GADV]-peptides were synthesized to maintain a high osmotic pressure, is the most essential matter for life, but not the genetic system.

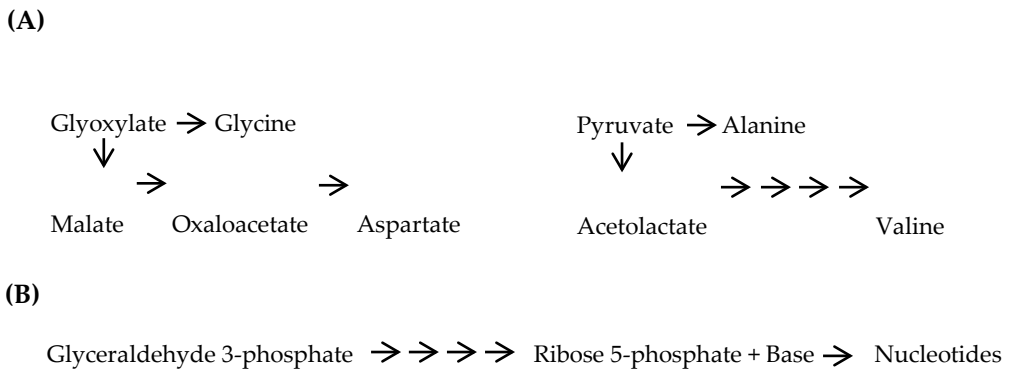


Figure 7. (A) Proto-metabolic pathways for [GADV]-amino acid synthesis. Four [GADV]-amino acids were produced from glyoxylate and pyruvate as starting materials through the proto-metabolic pathways. **(B)** Proto-metabolic pathways for nucleotide synthesis. Four nucleotides were produced with ribose 5-phosphate, which was synthesized from glyceraldehyde 3-phosphate as starting material through the proto-metabolic pathways.

As well known, modern cell membrane is composed of phospholipids and membrane proteins. Membrane proteins but not phospholipids exhibit various membrane functions. On the contrary, phospholipids are used for filling the gaps among membrane proteins and for expediting migration of membrane proteins to express membrane functions more efficiently. Therefore, it would be valid to consider that phospholipids were inserted into [GADV]-protein membrane after formation of phospholipid synthetic pathways through proto-metabolism in [GADV]-microsphere.

5. Formation of proto-metabolic pathways for nucleotide synthesis

To establish the core life system, sufficient amounts of nucleotides necessary to synthesize RNA must be produced with previously existing immature [GADV]-proteins. For the purpose, proto-metabolic pathways, through which nucleotides could be synthesized, must be formed (Figure 7 (B)). However, needless to state, formation of metabolic pathways in the absence of gene must rely on random processes.

Therefore, I consider that nucleotides must be produced through proto-metabolism using glyceraldehyde as a starting compound for ribose 5-phosphate synthesis [1; Chapter 5]. The reason, why nucletide synthetic pathway could be formed with immature

 doi:10.20944/preprints202206.0370.v1

[GADV]-proteins, is because, in addition to pluripotency of the immature [GADV]-proteins [1; Chapter 3], [GADV]-microspheres, which acquired more favorable metabolic pathways for proliferation than others, even accidentally, could leave more descendants than others. I would like to name the phenomenon as "proto-Darwin evolution". In this way, during the evolutionary process, metabolic pathways including nucleotide synthesis, which were favorable for [GADV]-microspheres to proliferate, were formed in the microspheres.

However, many researchers may consider that nucleotides never be synthesized from glyceraldehyde with immature [GADV]-proteins, if they do not know the significance of pluripotency of immature [GADV]-proteins which could be synthesized by direct random joining of [GADV]-amino acids under the protein 0^{th}-order structure [1; Chapter 3]. So, I would like to stress that the mystery of the origin of life never be solved as long as they rely the nucleotide synthesis only on the prebiotic means. Similarly, it would be reasonable to consider that the RNA world never be formed on the primitive Earth, because it would be impossible to produce a sufficient amount of RNA leading to the emergence of life without nucleotide metabolic pathway.

2. Formation of [GADV]-aa-AntiC-SL tRNA

1. Evolution of activated [GADV]-amino acids

Initially, synthesis of [GADV]-amino acids in [GADV]-microspheres was carried out by immature [GADV]-proteins, which were produced by direct joining of [GADV]-amino acids, such as through wet-drying processes in depressions of rocks on the primitive Earth. However, [GADV]-peptides could be produced at a much faster rate by using activated [GADV]-amino acids as [GADV]-AMP than the direct use of [GADV]-amino acids. After that, activated [GADV]-amino acids were used for more efficient synthesis of [GADV]-peptides in the following order, although there was no difference between direct joining of [GADV]-amino acids and the [GADV]-peptide synthesis with activated [GADV]-amino acids except difference of the reaction rate.

1-1. Use of [GADV]-aa-AMP: [GADV]-aa-AMPs were synthesized with immature [GADV]-proteins (actually [GADV]-peptide aggregates) to accelerate the peptide synthesis, after ATP was synthesized and accumulated at a large amount in [GADV]-microspheres. At this time point, it is supposed that the activated [GADV]-amino acids were exclusively used for the peptide synthesis with immature [GADV]-protiens owing to the accumulation of ATP in the microspheres.

1-2. Use of [GADV]-aa-3'-ACC: Successively, [GADV]-aa-3'-ACCs were used for the synthesis of [GADV]-peptides. Stability of trinucleotide, CCA-3', against RNase activity of immature [GADV]-proteins made it possible to use for the synthesis. It would be easily understandable that the single-stranded (ss)-trinucleotide 3'-ACC end of modern tRNA is also stable against RNase. Needless to state, the synthesis of [GADV]-peptides with activated [GADV]-amino acids were carried out non-specifically, because such activators, as

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 June 2022 doi:10.20944/preprints202206.0370.v1

AMP and 3′-ACC, cannot generate specificity to the respective [GADV]-amino acids. However, the use of such activators contributed to more efficient synthesis of [GADV]-peptides, because sites of 3′-ACC more than ATP, itself, could be used for binding with immature [GADV]-protein enzyme and the use of activated [GADV]-amino acids with a high energy became possible.

Of course, not only amino acids but also other organic compounds might be activated with ATP and 3′-ACC. In the case of the peptide synthesis with activated amino acids too, messy organic compounds should be incorporated into the peptides (Figure 4). However, it is supposed that activated organic compounds other than [GADV]-amino acids did not eventually contribute to the emergence of life, because functions of [GADV]-peptides containing meaningless organic compounds would be lowered.

1-3. Use of 3′-ACC-AntiC-SL RNA: After the use of 3′-ACC, [GADV]-peptide synthesis was carried out using 3′-ACC-AntiC-SL RNA (Figure 8). One of the reasons, why the AntiC-SL RNA was used for activation of [GADV]-amino acids, is because the AntiC-SL RNA, which were produced during cycles of oligonucleotide sysnthesis and degradation of the oligonucleotides, was the smallest but sufficiently stable RNA against hydrolysis by immature [GADV]-proteins. In addition, association of two AntiC-SL tRNAs side by side through base pairing between U and A in the two AntiC-loop RNAs facilitated the peptide bond formation between two amino acids bound to the 3′-ACC-end [1; Chapter 7].

The AntiC-SL tRNA primitive tRNA hypothesis is supported by the facts that base pairs between two complementary GNCs are stable [28] and also that any base in anticodon loop of three AntiC-SLs except Asp-tRNA out of four modern *Escherichia coli* [GADV]-aa-tRNAs is not chemically modified [29].

Part III. Steps from formation of AntiC-SL RNA to the emergence of life

1. Formation of ds-(GNC)$_n$ RNA gene

1. Establishment of the core life system in [GADV]-microsphere

Of course, the initial [GADV]-proteins, which were produced with immature [GADV]-proteins in a [GADV]-microsphere, were not literally [GADV]-proteins, in the meaning that [GADV]-proteins were contained non-natural amino acids other than [GADV]-amino acids. Synthesis of such impure polypeptides should always occur before formation of the first gene or in the absence of gene on the primitive Earth. However, even the impure and immature [GADV]-proteins could advanced catalytic reactions in the microspheres, although the activities were low. Therefore, literal [GADV]-proteins could not be produced until GNC primeval genetic code and (GNC)$_n$ gene were formed. Inversely stating this, selection of [GADV]-microsphere with more pure [GADV]-protein with a higher catalytic activity made it possible to form the first gene leading to synthesis of pure [GADV]-proteins. Consequently, the genetic system or the core life system composed of

 doi:10.20944/preprints202206.0370.v1

$(GNC)_n$ gene, AntiC-SL tRNA (GNC code) and [GADV]-protein, was invented to improve [GADV]-protein functions through complete exclusion of non-natural amino acids [1; Chapter 6]. Thus, the genetic system was formed to establish the most primitive but pure [GADV]-protein synthesizing system.

Needless to state, the first gene encoding the first mature [GADV]-protein with an ordered amino acid sequence must be generated through evolutionary process containing at least one random process as described below.

(1) Synthesis of a ss-$(GNC)_n$ RNA: In this case, the key point is to know the process how the first ss-$(GNC)_n$ RNA was formed, because an RNA with random $(GNC)_n$ codon sequence could be formed by random joining of GNC anticodons carried by the AntiC-SL RNAs. Note that the synthetic process of an immature [GADV]-protein with a random [GADV]-amino acid sequence through the random $(GNC)_n$ codon sequence is essentially the same with the [GADV]-protein synthesis by direct joining of [GADV]-amino acids. Then, how was the ss-$(GNC)_n$ RNA formed? I consider the formation process as follows.

1-1. Two pairs of two AntiC-SL tRNAs, which were bound in a column using two complementary GNC anticodons, were further aligned side by side to make tetramer of four AntiC-SL RNAs (Figure 8).

1-2. Two anticodons of the two AntiC-SL tRNAs, which were aligned side-by-side, were connected by phosphodiester bond. Thus, a random $(GNC)_n$ RNA sequence encoding random [GADV]-amino acid sequence was created (Figure 8).

However, many researchers may doubt whether the ss-$(GNC)_n$ RNA could be formed as described above. My answer to the question is as follows.

1-3. The consecutive codons on mRNA have been actually read by two anticodons of two adjacent modern tRNAs.

1-4. The serial reading mechanism of codon sequence on mRNA with anticodon of tRNA clearly indicates that it is reasonable from a stereochemical viewpoint to consider that a ss-codon sequence (mRNA) can be produced by joining of anticodons of tRNAs. That is the reason why successive codon sequence can be read by anticodons of two tRNAs tightly bound side-by-side. However, it would be difficult for two tRNAs to read a successive codon sequence, if the codon sequence was formed independently of tRNA. Genetic sequence on mRNA was translated by AntiC-SL tRNA as the reverse process of formation of $(GNC)_n$ codon sequence in [GADV]-microsphere. Thus, the fact that the commaless codon sequence can be translated by tRNAs indicates that genetic sequence was formed by random joining of anticodons carried by tRNAs.

Note that the $(GNC)_n$ codon sequence is the simplest but a meaningful sequence, which can be produced through a random process and also can be used for immature [GADV]-protein synthesis. This also indicates that formation of $(GNC)_n$ codon sequence by random joining of GNC anticodons of AntiC-SL tRNA is only one way, under which

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 June 2022 doi:10.20944/preprints202206.0370.v1

an immature but meaningful [GADV]-protein could be produced through the RNA sequence, which was formed by the essentially random process.

(2) Formation of a ds-$(GNC)_n$ RNA: Successively, a ds-$(GNC)_n$ RNA was formed by complementary strand synthesis of the ss-$(GNC)_n$ RNA.

(3) Formation of the first ds-$(GNC)_n$ RNA gene: Finally, the first $(GNC)_n$ RNA gene was formed. The formation of the first $(GNC)_n$ gene was triggered by synthesis of an immature [GADV]-protein from one strand of ds-RNA. The $(GNC)_n$ codon sequence encoding a random [GADV]-amino acid sequence evolved to one ds-RNA gene encoding a mature [GADV]-protein as led by promotion of the activity on the immature [GADV]-protein (Figures 2 and 8).

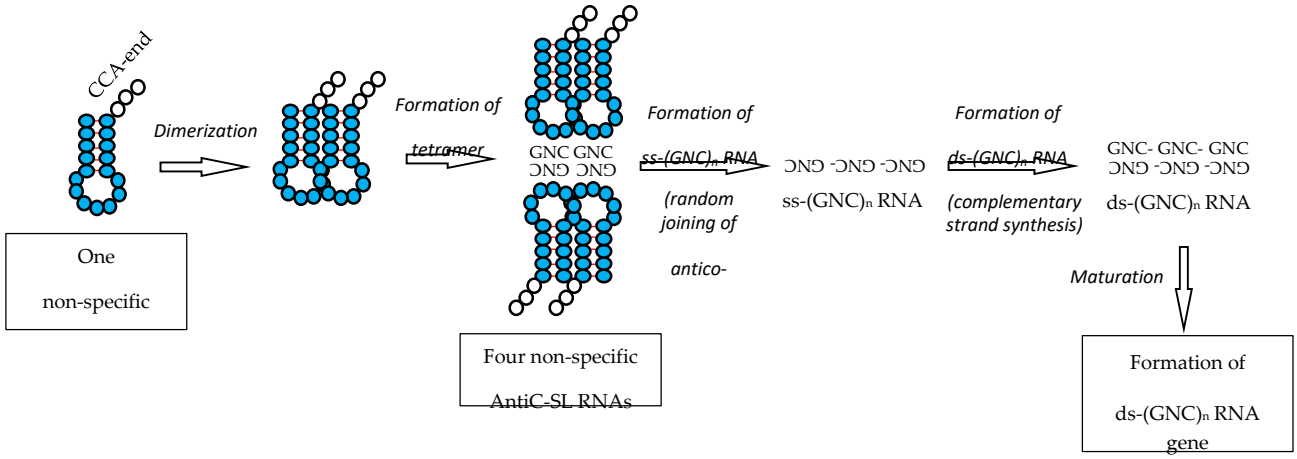


Figure 8. Formation process of ds-RNA gene. Ds-RNA gene was formed through ss-$(GNC)_n$ RNA and ds-$(GNC)_n$ RNA. The ss-$(GNC)_n$ RNA was formed by random joining of GNC anticodons carried by AntiC-SL RNAs. The use of amino acids were restricted into four [GADV]-amino acids accompanied by establishment of GNC primeval genetic code and ds-$(GNC)_n$ RNA gene for the first time.

Finally, ds-$(GNC)_n$ RNA genes encoding a mature [GADV]-protein were formed as the most effective means for producing mature [GADV]-proteins with a large catalytic activity. It should be taken notice again that the steps towards formation of the first gene were the process for selecting only [GADV]-amino acids more efficiently from chemically complex messy environments. Thus, the key of the first ds-$(GNC)_n$ gene formation was

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 June 2022 doi:10.20944/preprints202206.0370.v1

the formation of a ss-$(GNC)_n$ RNA with random GNC codon sequence through a random process.

Thus, the synthesis of mature [GADV]-proteins composed of only [GADV]-amino acids became possible for the first time after ds-$(GNC)_n$ RNA gene was acquired and the translation system using four kinds of [GADV]-aa-tRNAs were established. Note that all members, which need to execute both mRNA synthesis and translational process, had completed already at the time point when the first ds-RNA gene was created, because both the use of transcription and translation systems had already become possible, just after not the ds-$(GNC)_n$ RNA gene but the ds-$(GNC)_n$ RNA was formed [1; Chapter 6].

6. The emergence of life

(1) Genuine lives arose after acquisition of various $(GNC)_n$ genes through creation of homologous genes and entirely new $(GNC)_n$ genes, which were derived from sense strands [30] and antisense strands [15] of previously established $(GNC)_n$ genes, respectively [1; Chapter 8]].

(2) The first life emerged not at one moment but during consecutive changes. In other words, any critical moment of the emergence of the first life did not exist. The emergence of life would be such a gradual change as it can be confirmed that a typical life had arisen after a time point.

Discussion

It would be difficult to give an answer to the question how life emerged in "chemically complex messy environments". The problem is essentially the same question as how the most primitive core life system, which is composed of [GADV]-protein, AntiC-SL tRNA (GNC primeval genetic code) and $(GNC)_n$ RNA gene using the respective small number of monomers, as four [GADV]-amino acids and four nucleotides, was established in the messy environments, because an answer to the mystery of the origin of life is, simultaneously, the answer to the question how the small number of monomers were selected out from the chemically complex messy environments on the primitive Earth. Therefore, it would be necessary to consider as focusing on the discussion how life arose using the core life system on the primitive Earth in order to find the answer to the question how life emerged in "chemically complex messy environments.

The answer to the question is given based on GADV hypothesis on the origin of life by considering several factors comprehensively, as described below.

1. A small kinds of, at least not so many organic compounds including [GADV]-amino acids could be synthesized from inorganic compounds with prebiotic means on the primitive Earth.

 doi:10.20944/preprints202206.0370.v1

2. [GADV]-polypeptides composed of mainly four [GADV]-amino acids could be produced through a random process on the primitive Earth. The reason is because [GADV]-amino acids having both positive and negative charges in the molecule pulled against each other owing to the electrostatic attraction. Immature [GADV]-proteins, actually aggregates of [GADV]-peptides, which were produced by joining of [GADV]-amino acids randomly selected from [GADV]-amino acid composition or one of the protein 0^{th}-order structures, could be folded into water-soluble globular structure to exhibit a weak but effective various functions [26].

3. [GADV]-microspheres could be formed with immature [GADV]-proteins, actually [GADV]-peptide aggregates. Formation of the [GADV]-microspheres, which demonstrate individuality, made it possible to grow, proliferation and evolution.

4. Furthermore, metabolic pathways for [GADV]-amino acid synthesis using simple organic compounds as glyoxylate and pyruvate, which accumulated on the primitive Earth, could be formed using immature [GADV]-proteins in the microspheres ([GADV]-protein world). The formation of the [GADV]-amino acid metabolic pathways assured to continuously grow, divide and proliferate [GADV]-microspheres through synthesis of [GADV]-peptides.

It may become a matter whether or not [GADV]-amino acids could be synthesized with immature [GADV]-proteins, using as starting compounds, glyoxylate and pyruvate. However, [GADV]-amino acid synthetic pathways could be formed if such immature [GADV]-proteins, which have various catalytic activities or pluripotency as catalysts, were used as biocatalysts [1; Chapter 3,26,27].

5. Metabolic pathway for synthesis of four nucleotides could be also formed using immature [GADV]-proteins in [GADV]-microspheres, similarly as the case of synthetic pathways of [GADV]-amino acids.

6. Further, it is considered that AntiC-SL tRNAs were formed with four kinds of nucleotides synthesized through the metabolic pathways. Thus, the use of nucleotides were restricted for the first time into four, or adenine (A), guanine (G), uracil (U) and cytosine (C), because two base pairs, A-U and G-C, was indispensable to fold RNA strand into the AntiC-SL through hydrogen bonds with a large directionality.

7. Then, GNC primeval genetic code, which determines the framework composed of four [GADV]-amino acids and four GNC codons, was established although the corresponding relationships between GNC codons and [GADV]-amino acids were accidentally determined and frozen as assumed by the GNC code frozen-accident theory [1; Chapter 7].

8. Eventually, the first genuine life emerged on the primitive Earth accompanied, after (GNC)n RNA genes were formed successively in order of ss-$(GNC)_n$ RNA, ds-$(GNC)_n$ RNA and ds-$(GNC)_n$ RNA gene.

Thus, the first genuine life using biopolymers composed of a small kinds of components emerged in the chemically complex messy environments on the primitive Earth about 4 billion years ago. I believe that such steps to the emergence of life was inevitable and there was no way to the emergence of life. In this way, it can be reasonably explained based on the GADV hypothesis on the origin of life how the core life system composed of $(GNC)_n$ gene, AntiC-SL tRNA and [GADV]-proteins, all of which are composed of a small kinds of components, could be established in messy environments on the primitive Earth. This indicates that GADV hypothesis is a valid idea for explaining the steps to the emergence of life.

Acknowledgements

I am very grateful to Dr. Tadashi Oishi (G&L Kyosei Institute, Emeritus professor of Nara Women's University) for encouragement throughout my research on the origin and evolution of the fundamental life system .

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ikehara, K.; *Towards revealing the origin of life. -Presenting the GADV hypothesis-;* (Monograph) Springer Nature, Gewerbestrasse Cham, Switzerland **2021**.
2. Gilbert, W.; The RNA world. *Nature*, **1986**, 319, 618.
3. Ikehara, K.; Origins of gene, genetic code, protein and life: Comprehensive view of life system from a GNC-SNS primitive genetic code hypothesis. *J. Biosci*., **2002**, 27:165–186.
4. Ikehara, K.; Possible steps to the emergence of life: The [GADV]-protein world hypothesis. *Chem. Rec*., **2005**, 5, 107–118.
5. Ikehara, K.; Pseudo-replication of [GADV]-proteins and origin of life. *Int. J. Mol. Sci*., **2009**, 10, 1525–1537.
6. Dill, K.A.; Dominant forces in protein folding. *Biochemistry*, **1990**, 297, 133–7155.
7. Kim, S.C.; O'Flaherty, D.K.; Giurgiu, C.; Zhou, L.; Szostak, J.W.; The emergence of RNA from the heterogeneous products of prebiotic nucleotide synthesis. *J. Am. Chem. Soc*., **2021**, 143(9), 3267-3279. doi: 10.1021/jacs.0c12955
8. Kim, S.C.; Zhou, L.; Zhang, W.; O'Flaherty, D.K.; Rondo-Brovetto, V.; Szostak, J.W.; A Model for the Emergence of RNA from a Prebiotically Plausible Mixture of Ribonucleotides, Arabinonucleotides, and 2'-Deoxynucleotides.

J. Am. Chem. Soc. **2020**, 142(5), 2317-2326. doi: 10.1021/jacs.9b11239.

9. Banfalvi, G.; Prebiotic Pathway from Ribose to RNA Formation. *Int. J. Mol. Sci*., **2021**, 22(8), 3857. doi: 10.3390/ijms22083857.
10. Benner, S.A.; Kim, H.J.; Carrigan, M.A.; Asphalt, water, and the prebiotic synthesis of ribose, ribonucleosides, and RNA. *Acc. Chem. Res*., **2012**, 45(12), 2025-34. doi: 10.1021/ar200332w.
11. Orgel, L.E.; Prebiotic chemistry and the origin of the RNA world. *Crit. Rev. Biochem. Mol. Bio*l., **2004**, 39(2), 99-123. doi: 10.1080/10409230490460765.
12. Unrau, P. J.; Bartel, D.P.; RNA-catalysed nucleotide synthesis. *Nature*, **1998**, 395(6699), 260-3. doi: 10.1038/26193.
13. Ikehara, K.; Evolutionary Steps in the Emergence of Life Deduced from the Bottom-Up Approach and GADV Hypothesis (Top-Down Approach). *Life (Basel)*, **2016**, 6, 6.
14. Ikehara, K. Protein ordered sequences are formed by random joining of amino acids in protein 0^{th}-order structure, followed by evolutionary process. *Orig. Life Evol. Biosph.*, **2014**, 44, 279–281.
15. Ikehara, K.; Amada, F.; Yoshida, S.; Mikata, Y.; Tanaka, A.; A possible origin of newly-born bacterial genes: significance of GC-rich nonstop frame on antisense strand. *Nucl. Acids Res.*, **1996**, 24, 4249-425.
16. Miller, S.L.; Orgel, L.E.; *The origins of life on the earth*; Prentice-Hall: Englewood Cliffs, NJ, USA. **1974**.
17. Cleaves, H.J., Chalmers, J.H.; Lazcano, A.; Miller, S.L.; Bada, J.L.; A reassessment of prebiotic organic synthesis in neutral planetary atmosphere. *Orig. Life Evol. Biosph*., **2008**,, 38:105–115.
18. Clay, A.P.; Cooke, R.E.; Kumar, R.; Yadav, M.; Krishnamurthy, R.; Springsteen, G.; A plausible prebiotic one-pot synthesis of orotate and pyruvate suggestive of common protometabolic pathways. *Angew. Chem. Int. Ed. Engl*., **2022** , 61(11): doi: 10.1002/anie.202112572.
19. Zhou, R.; Basu, K.; Hartman, H.; Matocha, C.J.; Sears, S.K.; Vali, H.; Guzman, M.I.; Catalyzed synthesis of zinc clays by prebiotic central metabolites. *Sci Rep*. **2017**, 7(1): 533. doi: 10.1038/s41598-017-00558-1.
20. Schwartz, A.W.; Chittenden, G.J.; Synthesis of uracil and thymine under simulated prebiotic conditions. *Biosystems*, **1977**, 9(2-3): 87-92. doi: 10.1016/0303-2647(77)90016-8.
21. Pearce, B.K.D.; Tupper, A.S.; Pudritz, R.E,; Higgs, P.G.; Constraining the Time Interval for the Origin of Life on Earth. *Astrobiology*, **2018**, 18(3): 343-364. doi: 10.1089/ast.2017.1674.
22. Ehrenfreund, P.; Spaans, M.; Holm, N.G.; The evolution of organic matter in space. *Philos Trans A Math Phys Eng Sci*. **2011**, 369(1936): 538-54. doi: 10.1098/rsta.2010.0231.PMID: 21220279
23. Pizzarello, S.; Weber, A.L.; Prebiotic amino acids as asymmetric catalysis. *Science*, **2004**, 303:1151.
24. Furukawa Y, Chikaraishi Y, Ohkouchi N.; Ogawa, N.O.; Glavin, D.P.; Dworkin, J.P.; Abe, C.; Nakamura, T.;

Extraterrestrial ribose and other sugars in primitive meteorites. Proc Natl Acad Sci., **2019**,116:24440-24445

25. Oba, T.; Fukushima, J.; Maruyama, M.;. Iwamoto, R.; Ikehara, K.; Catalytic activities of [GADV]-peptides: Formation and establishment of [GADV]-protein world for the emergence of life. *Orig. Life Evol. Biosph.* **2005**, 35, 447-460.
26. Luisi, P.L.; "*The emergence of life-from chemical origins to synthetic biology*- 2nd ed.", Cambridge Univ. Press, Cambridge, United Kingdom, **2016**.
27. Ikehara, K.; *GADV hypothesis on the origin of life -Life emerged in this way!?* LAP LAMBERT Academic Publishing, Saarbrucken, Germany, **2016**.
28. Taghavi, A.; van der Schoot, P., Berryman, J.T.; DNA partitions into triplets under tension in the presence of organic cations, with sequence evolutionary age predicting the stability of the triplet phase. *Q. Rev. Biophys*., **2017**, e15 https://doi.org/10.1017/S0033583517000130
29. tRNAdb, transfer RNA database (Universitat Leipzig): http//trnadb.bioinf.uni-leipzig.de.
30. Ohno, S.; *Evolution by gene duplication*, Springer, Heiderberg, Germany, **1970**.