

Review

Not peer-reviewed version

Solving the Mystery of Origin of Life— GADV Hypothesis

[Kenji Ikehara](#) *

Posted Date: 11 December 2024

doi: 10.20944/preprints202412.0861.v1

Keywords: GADV hypothesis; GADV protein world; origin of life; origin of gene; origin of genetic system



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Review

Solving the Mystery of Origin of Life—GADV Hypothesis

Kenji Ikehara ^{1,2}

¹ G&L Kyosei Institute, In Keihanna Academy of Science and Culture (KASC), Keihanna interaction plaza, Lab. Wing 3F, 1-7 Hikaridai, Seika-cho, Souraku, Kyoto 619-0237, Japan; ikehara@cc.nara-wu.ac.jp; Tel.: +81-774-73-4478

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

Abstract: The origin of life has been mainly studied based on three ideas, RNA world hypothesis, hydrothermal vent hypothesis and space-origin hypothesis. However, the “mystery” has not been solved still now in spite of strenuous efforts of many researchers for more than 40 years. On the contrary, I proposed one idea, [GADV]-protein world hypothesis (GADV hypothesis), about 20 years ago, which insists that life emerged from [GADV]-protein world formed by pseudo-replication of [GADV]-proteins. [GADV] or GADV means four amino acids, Gly [G], Ala [A], Asp [D] and Val [V]. It was found in subsequent studies that the first gene was formed through processes starting with immature [GADV]-proteins or in a bottom-up manner. Furthermore, it has been confirmed that the first genetic system composed of ds-(GNC)_n RNA gene, anticodon stem-loop (AntiC-SL) tRNA (GNC primeval genetic code) and [GADV]-protein was established at the same time when the first gene was formed. It has been also confirmed that the genetic system never be established through a process starting with gene or tRNA. From the results, it has been concluded that life arose as expected by GADV hypothesis and that the “mystery” of the origin of life has been solved.

Keywords: GADV hypothesis; GADV protein world; origin of life; origin of gene; origin of genetic system

1. Introduction

The studies on the origin of life have been carried out under the following three major hypotheses, RNA world hypothesis [1–3], hydrothermal vent hypothesis [4–6] and space-origin hypothesis [7–9]. However, the “mystery” of the origin of life has not been solved in spite of strenuous efforts by many researchers for several decades. The main reasons would be as follows. (1) Many researchers have overlooked an important concept, which are necessary to solve the “mystery”, that is, protein 0th-order structure, which is crucial to generate an immature but meaningful water-soluble globular protein under a random process in the absence of gene or on the primitive Earth [10]. (2) The first genetic code, GNC primeval genetic code, has not been sufficiently understood by many researchers. For the reasons, formation processes of both the genetic system and reasonable steps to the emergence of life have not been explained. Another reason would be because studies on the origin of life were carried out with a focus on RNA world hypothesis, in which the origin of life has been studied in top-down manner starting with gene [11].

On the contrary, I have proposed [GADV]-protein world hypothesis (GADV hypothesis) [12,13], with which I consider that the “mystery” of the origin of life might be solved, as described in this review article. Then, it is first described in this article, through what processes I did propose the GADV hypothesis, because notice of the hypothesis was triggered by a study on how entirely new genes are created in modern microorganisms [14,15], which is apparently irrelevant to the origin of life (Figure 1). Inversely stating this, that would be one of the reasons why almost all researchers working in the research field of the origin of life could not notice the protein 0th-order structure or

[GADV]-amino acid composition [10] and the GADV hypothesis [12,13]. [GADV] and GADV mean four amino acids, Gly [G], Ala [A], Asp [D] and Val [V].

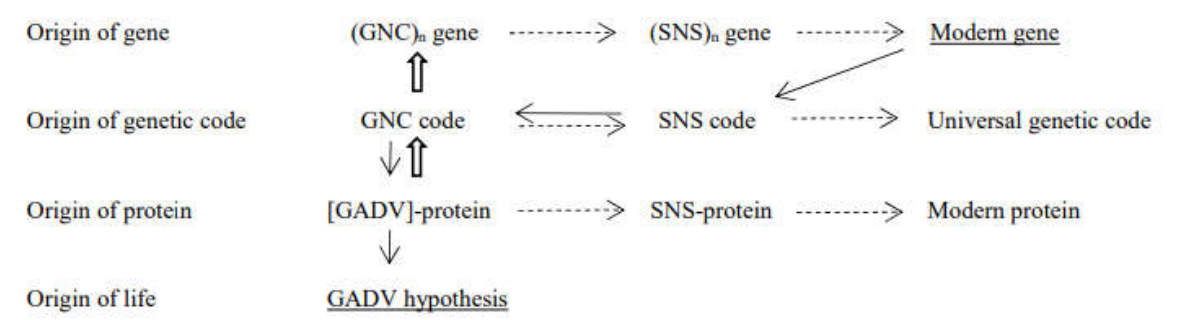


Figure 1. Research process, through which GADV hypothesis could be hit upon. The research pursuing the GADV hypothesis started from the origin of modern gene indicated by underline. Thereafter, GNC primeval genetic code was proposed via SNS primitive genetic code [16,17]. Finally, GADV hypothesis on the origin of life shown by underline could be presented [12,13], which is based on pseudo-replication of [GADV]-proteins [18]. The research progresses are shown by thin arrows. Formation processes of the first (GNC)_n gene and evolutionary processes of gene, genetic code and protein are indicated by bold white arrows and broken arrows, respectively.

2. Processes, Which I Fortunately Came Across the GADV Hypothesis

In this section, the processes, which I hit upon GADV hypothesis, are explained, because the reason, why only I noticed the hypothesis, could be understood.

2.1. Exploration of Origin of Modern Gene

The study of the GADV hypothesis started about 35 years ago from the study for solving a question how and where an entirely new gene, which is quite different from any gene existing at that time, is generated, if such entirely new genes are generated on this planet still now. Consequently, GC-NSF(a) hypothesis suggesting that entirely new genes encoding an entirely new protein are generated from nonstop frame on antisense strand of GC-rich genes (GC-NSF(a)) was proposed [14,15].

2.2. Discovery of SNS Primitive Genetic Code

Successively, I examined on general features of the GC-NSF(a)s in order to understand the reason why entirely new genes can be formed from the GC-NSF(a)s. Consequently, it was found that base compositions at three base positions of the GC-NSF(a) have nearly SNS pattern or (G/C)N(G/C) [15]. This implies that water-soluble globular proteins could be produced with ten types of amino acids encoded by the hypothetical SNS code (Figure 1) [16].

2.3. Discovery of GNC Primeval Genetic Code

Furthermore, more simple and primitive genetic code was explored. From the results, GNC code, which is composed of four GNC codons and four [GADV]-amino acids, was obtained as the oldest genetic code [17]. Based on the results, GNC-SNS primitive genetic code hypothesis suggesting that the universal genetic code has been derived from GNC code through SNS code, was proposed about 20 years ago (Figure 1) [17].

2.4. Proposition of GADV Hypothesis

Finally, an idea, that the first life arose from [GADV]-protein world, which was formed by pseudo-replication of [GADV]-proteins [18] under a protein 0th-order structure [10], suddenly hit upon in my head. The idea was presented as [GADV]-protein world hypothesis or GADV hypothesis [12,13] (Figure 2). The protein 0th-order structure is a special amino acid composition containing

[GADV]-amino acids at roughly equal amounts [10]. Owing to one of the protein 0th-order structures, immature but sufficiently stable water-soluble globular proteins with some flexibility could be produced even by direct random joining of four [GADV]-amino acids in the absence of gene, or even before the first gene could be formed.

At that time, only the steps from formation of immature [GADV]-proteins to the emergence of life were vaguely appeared in my head (Figure 2) [13].

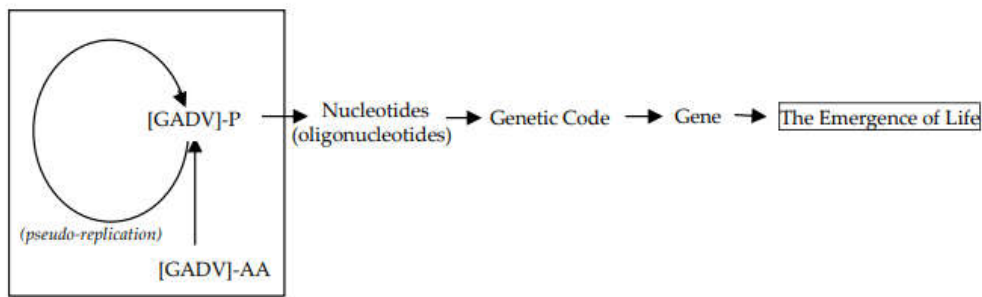


Figure 2. [GADV]-protein world hypothesis (GADV hypothesis) [12,13], which was firstly proposed as a starting point of GNC-SNS primitive genetic code hypothesis [16]. Therefore, only three members, protein, genetic code and gene, are appeared on the way to the emergence of life [13].

3. Six Steps to the Emergence of Life

Thereafter, more detailed steps to the emergence of life were explored through investigation of indispensable matters for modern life to live [19], as described below.

3.1. Indispensable Matters for Modern Life to Live

- (1) As a matter of course, the genetic system composed of four members, gene, tRNA (genetic code) and protein is indispensable.
- (2) Metabolic system, which is carried out by proteins produced through the genetic system, is also indispensable.
- (3) Furthermore, cell structure enclosing the genetic and the metabolic systems is indispensable.

Therefore, these mean that it is necessary to make clear the respective formation processes of the six members in order to elucidate the origin of life. New evolutionary process from chemical evolution to the emergence of life, which has been revised by introduction of [GADV]-microsphere (cell structure), proto-metabolic system (metabolism) and anticodon stem-loop (AntiC-SL) RNA (tRNA) into the three steps shown in Figure 2, is given in Figure 3 [19].

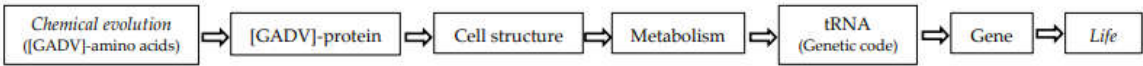


Figure 3. The steps, which are deduced by GADV hypothesis from chemical evolution to the emergence of life. A scenario to the emergence of life is given in the figure based on the origins of six members composing the fundamental life system [19].

Then, explain the formation processes of the respective six members (Figure 4) [19].
Step 1: Production of immature [GADV]-protein
[GADV]-peptides were synthesized by random joining of [GADV]-amino acids, which accumulated on the primitive Earth, for example in depressions of rocks on seashore of the primitive Earth by repeated wet-drying cycles. Successively, immature but water-soluble globular [GADV]-proteins with some flexibility were formed by association of [GADV]-peptides having a random [GADV]-amino acid sequence.
Furthermore, [GADV]-peptides were synthesized by immature [GADV]-proteins with peptide bond forming catalytic activity or by pseudo-replication [18].

Step 2: Formation of [GADV]-microspheres

[GADV]-microspheres were formed with immature [GADV]-proteins (actually [GADV]-peptide aggregates) [19].

Step 3: Formation of primeval metabolic system

Next, a primeval metabolic system was formed in [GADV]-microspheres by immature but pluripotent [GADV]-proteins using three organic compounds, glyoxylate (Go), glyceraldehyde (Ga) and pyruvate (Pyr), which were incorporated from outside of the microspheres [19].

Step 4: Formation of AntiC-SL RNA (proto-tRNA)

ATP was synthesized by bonding ribose, which was synthesized from glyceraldehyde, with adenine, which was incorporated from outside of the cell structure. ATP was used as an activator of [GADV]-amino acids for peptide synthesis. Furthermore, four nucleotides, which were synthesized by immature [GADV]-proteins, accumulated gradually in [GADV]-microsphere. Proto-tRNA (AntiC-SL RNA) or one of activators for [GADV]-protein synthesis was formed through repeated random joining of nucleotides and degradation of the oligonucleotides with immature [GADV]-proteins in [GADV]-microspheres [19,20]. Note that AntiC-SL RNA was the smallest but sufficiently stable RNA composed of 17 nucleotides having a activation function of [GADV]-amino acids.

Step 5: Establishment of primeval GNC genetic code

The origin of genetic code has been considered thus far with a focus on the stereochemical theory [21,22]. However, formation process of the genetic code cannot be reasonably explained by the stereochemical theory, because trinucleotides, GNCs, are too large to form a stable complex with one of [GADV]-amino acids. Moreover, it would be impossible to use an amino acid in the complexes for peptide bond formation in all combinations, even if the complexes between an anticodon or a trinucleotide and the corresponding amino acid could be formed [19,23]. Therefore, a novel idea or the GNC-code frozen-accident theory has been proposed for explaining the establishment process of the first genetic code [19,23]. In other words, it is considered that the first GNC genetic code or the relationship between GNC anticodons and the corresponding [GADV]-amino acids was accidentally formed and was frozen.

Step 6: Formation of the first gene

(1) Formation of double-stranded (ds)-(GNC)_n RNA (proto-gene)

Single-stranded (ss)-(GNC)_n RNA was formed by random joining of anticodon GNCs carried by AntiC-SL RNA tetramers [19]. A ds-(GNC)_n RNA was thereafter formed through complementary strand synthesis of the ss-(GNC)_n RNA.

(2) Formation of the first ds-(GNC)_n RNA gene

The first gene encoding a mature [GADV]-protein was formed by maturation of an immature [GADV]-protein, which was produced by expression of either one of ds-(GNC)_n RNA strands (Figure 5) [19].

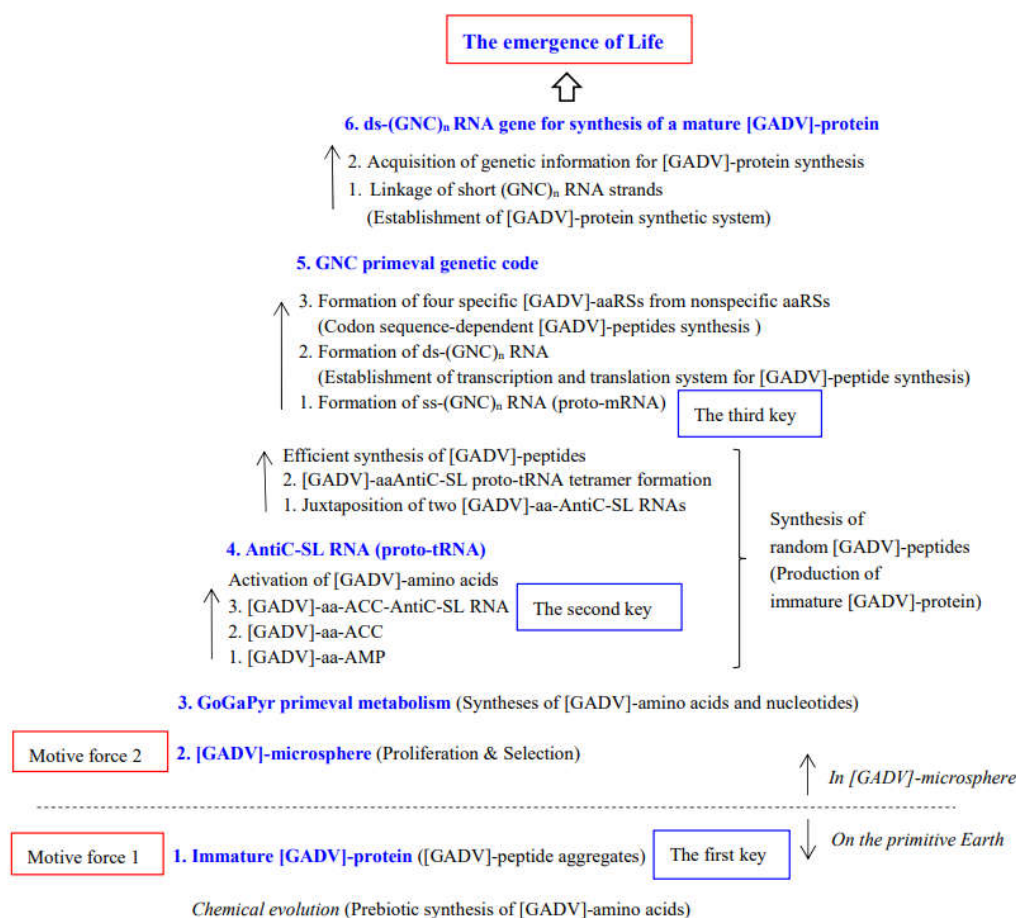


Figure 4. The overview of evolutionary process from chemical evolution to the emergence of life. In this figure, for example, the changes of activators for [GADV]-peptide synthesis and the role of aminoacyl-tRNA synthetase in establishment of GNC primeval genetic code are added to Figure 3.

Finally, the first life arose on the primitive Earth about 4 billion years ago after [GADV]-microsphere acquired a number of genes, which were necessary for the first life to live [19]. The steps from chemical evolution to the emergence of life are drawn in more detail in Figure 4.

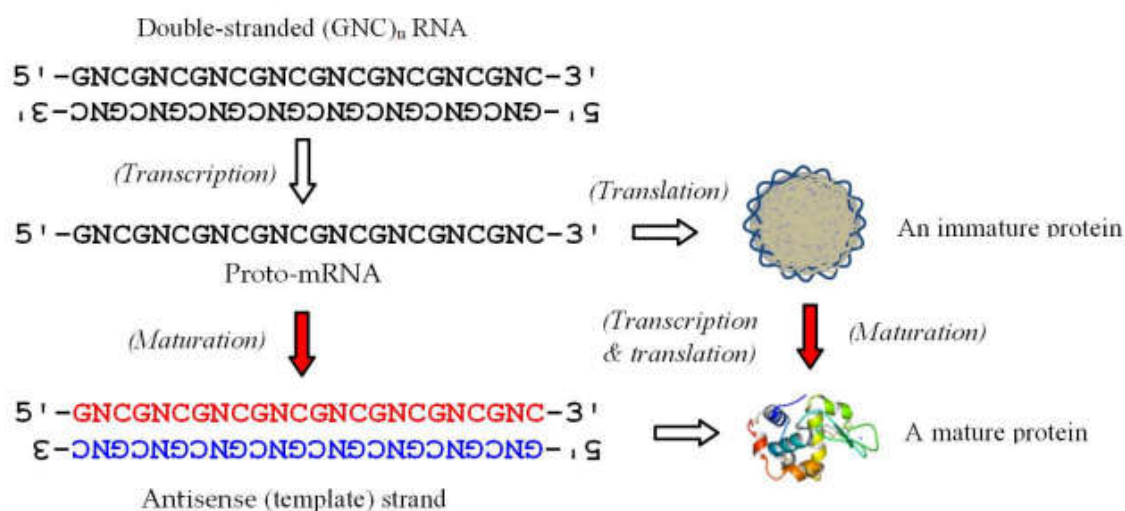


Figure 5. ds-(GNC)_n RNA must be first formed by complementary strand synthesis of ss-(GNC)_n RNA, which was synthesized by random joining of anticodons, GNC, carried by tetramer AntiC-SL RNAs (proto-tRNA) [11,19]. Successively, it is expected that the first gene encoding a mature protein was

formed through maturation of an immature protein, which was produced by gene expression of proto-mRNA. Red letters and blue letters indicate (GNC)_n codon sequence encoding a mature protein and (GNC)_n sequence on antisense strand, respectively.

4. Matters Indispensable to Solving the Mystery of Origin of Life

The overview of evolutionary process from chemical evolution to the emergence of life is shown in Figure 4. In this section it is confirmed whether or not the mystery of the origin of life could be solved under the GADV hypothesis. For the purpose, it is first enumerated what matters must be made clear to solve the mystery.

- (1) Formation process of the fundamental life system composed of six members
- (2) Formation process of the first gene
- (3) Formation process of the first genetic system

Then, the three formation processes are discussed below in order.

4.1. Formation Process of the Fundamental Life System

The fundamental life system of the first life should be also composed of the six members (protein, cell structure, metabolism, tRNA, genetic code and gene). Therefore, the respective origins of the six members must be not individually but comprehensively explained, because the six members should be intimately related each other even in the first life. Certainly, the formation processes could be explained as a series of formation processes of the members in GADV hypothesis (Figures 3 and 4). Thus, the first gene was formed after formation of the five members except gene. This means that the first gene would be formed as going upstream against the flow of genetic expression. The steps from chemical evolution to the emergence of life must be correct, because it has been confirmed that the order of any one of the six members cannot be exchanged for another member [19].

4.2. Formation Process of the First Gene Except Four [GADV]-Aminoacyl tRNA Synthetase (aaRS) Genes

Needless to state, the most important matter for the emergence of life is to acquire genetic information for formation of a mature protein. However, the acquisition process of genetic information could not be explained by any theory other than GADV hypothesis (RNA world hypothesis [1–3], Hydrothermal vent hypothesis [4–6], space origin hypothesis [7–9]) on the origin of life. The reason would be because the importance of protein 0th-order structure could not be well recognized.

Then, the matters, which played important roles in formation of the first gene, are enumerated as follows.

- (1) Double-stranded (ds)-RNA

ds-RNA is indispensable in order to write genetic information for synthesis of an immature [GADV]-protein under the protein 0th-order structure into either one of the double-strands (Figure 5).

- (2) RNA-dependent RNA polymerase

RNA polymerase is also necessary to express the genetic information on ds-RNA (Figure 5). It is expected that RNA polymerase activity of immature [GADV]-proteins could be used for the synthesis of a proto-mRNA, because it was reported by van der Gulik *et al.* that active site of RNA polymerase is mainly composed of [GADV]-amino acids [24].

- (3) Four specific proto-[GADV]-tRNAs

Four AntiC-SL tRNAs are necessary to translate GNC codon sequence on the ss-RNA. Such four AntiC-SL tRNAs carrying one of GNC anticodons could be formed through one base substitution at the second codon position of GNC carried by loop of the first AntiC-SL tRNA. The four GNC anticodons could be selected by formation of stable complementary GNC pairs carried by two AntiC-SL tRNAs, because it was already confirmed by Taghavi *et al.* that binding abilities between two complementary GNC base-pairs are stronger than other triplet base-pairs [25].

- (4) Four specific aaRSs

[GADV]-aminoacyl-tRNA synthetase ([GADV]-aaRS) should assist to combine one of [GADV]-amino acids with CCA end of the corresponding AntiC-SL RNA. Therefore, four genes encoding four specific [GADV]-aaRSs are also indispensable to synthesize the respective specific aaRSs. This means that there exists another “chicken-egg relationship” between the gene and the specific aaRS, because gene encoding a specific aaRS could not be expressed without the four specific aaRSs. On the contrary, synthesis of four specific [GADV]-aaRSs requires four genes encoding the respective specific aaRSs.

However, formation process of four specific [GADV]-aaRSs from four nonspecific aaRSs can be reasonably explained as follows.

1. ss-(GNC)_n RNAs, which were formed by random joining of GNC anticodons in four AntiC-SL RNAs.

2. ds-(GNC)_n RNA was formed by complementary strand synthesis of the ss-(GNC)_n RNAs. Even then, only nonspecific [GADV]-aaRSs could be produced by either strand of the ds-(GNC)_n RNAs. However, the four nonspecific [GADV]-aaRSs had the respective random but unique [GADV]-amino acid sequences, because the respective ds-(GNC)_n RNAs had a random but unique (GNC)_n codon sequence.

3. The respective [GADV]-aaRSs, which had a random but unique [GADV]-amino acid sequence, expressed slightly different or slightly specific activities against the respective [GADV]-amino acids and AntiC-SL RNAs.

4. The slightly specific activities of [GADV]-aaRSs should generate growth rate differences among [GADV]-microspheres.

5. [GADV]-microspheres having a larger growth rate caused by a specificity higher than others could be selected.

6. [GADV]-microspheres, in which four genes encoding four specific [GADV]-aaRSs could be fostered, were finally selected. Consequently, the four genes encoding four specific [GADV]-aaRSs were formed in the [GADV]-microspheres. Therefore, it is supposed that the first genes encoding a protein were such [GADV]-aaRS genes.

7. An immature [GADV]-protein with a unique [GADV]-amino acid sequence, which was produced from one of ds-(GNC)_n RNA and was matured under the specific four [GADV]-aaRSs.

Note that all of the events proceeded in [GADV]-microspheres slowly but steadily through selection of [GADV]-microspheres, in which immature [GADV]-aaRSs with a higher catalytic activity could be produced more efficiently than before. Consequently, four ds-(GNC)_n genes encoding a mature [GADV]-aaRS were formed (Figure 4).

Then, explain the acquisition process of the first genetic information for synthesis of a mature protein. Immature [GADV]-protein was synthesized by expression of either strand of the ds-(GNC)_n RNA with RNA polymerase activity of immature but pluripotent [GADV]-proteins, four specific [GADV]-AntiC-SL tRNAs and four specific [GADV]-aaRSs. Although it is needless to state, ability of ds-RNA memorizing base substitutions made it possible to evolve an immature [GADV]-protein to the mature protein (Figure 5) [19].

ds-(GNC)_n RNA was formed through from step 1 to step 5 as drawn in Figure 4. Once again, the reason, why the five steps could proceed from immature [GADV]-protein synthesis toward formation of ds-(GNC)_n RNA synthesis, is because [GADV]-microspheres with a system, in which more active immature [GADV]-proteins could be more efficiently produced than others, were selected. This means that immature [GADV]-proteins could be synthesized at the every step from step 1 to step 5 and that the steps proceeded as piling up on the first step (immature [GADV]-protein produced by direct random joining of [GADV]-amino acids) from the step 2 (formation of [GADV]-microsphere) to the step 5 (formation of ds-(GNC)_n RNA). In other words, formation processes to the emergence of life progressed as going upstream against the flow of gene expression or in a bottom-up manner (Figure 6).

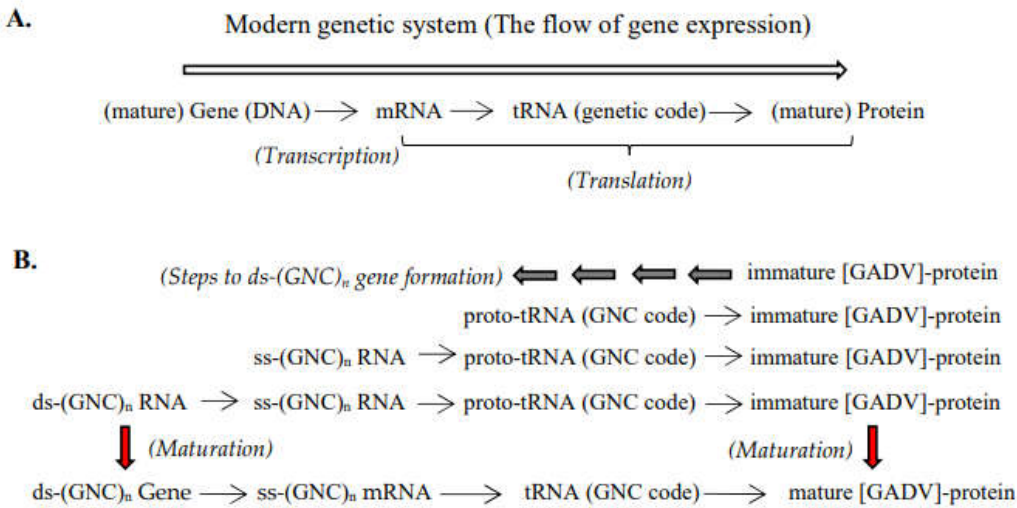


Figure 6. (A) The flow of gene expression in modern genetic system. Gene is expressed through two processes, transcription and translation. **(B)** It is considered in the GADV hypothesis that the first genetic system was formed from immature [GADV]-protein as going upstream against the flow of gene expression step by step (gray bold arrows). Note that immature [GADV]-protein is synthesized with each prototype at the respective evolutionary steps as shown by thin arrows. The reason, why the first gene and the first genetic system were formed in parallel, is because the first genetic system had established a the moment, when the first gene was formed through maturation of immature [GADV]-protein (red bold arrows).

Accompanied by formation of the first gene (step 6), both homologous genes and entirely new genes could be generated from a codon sequence on sense strand and an antisense strand of the first gene, respectively. The first genuine life arose after a number of genes necessary to live were equipped.

4.3. Formation Process of the First Genetic System

Formation process of the first genetic system, which is composed of gene, tRNA (genetic code) and protein, can be explained using the four members necessary to acquire genetic information (Figure 6 (A)). However, as a matte of course, writing process of genetic information into ds-(GNC)_n RNA is quite different from acquisition process of genetic expression system, because the first genetic information must be formed from protein or in the bottom-up manner and, contrary to that, genetic information flows from gene to protein inversely in the genetic system. Then, how was the first genetic system could be acquired?

It is described in the subsection 4.2 about the way how the first genetic information was written into either one of a ds-(GNC)_n RNA as going upstream against the flow of gene expression or in a bottom-up manner. However, the reason, why the first gene must be formed in the bottom-up manner, has not been well considered at that time. Therefore, the way how and the reason why the first genetic system was formed, has not been considered. However, when I had considered the problem for a while, it was conceived that the first genetic system, too, was formed and must be formed in the bottom-up manner in parallel with the first gene formation. The reason can be explained as follows.

(1) The first gene was formed owing to two motive forces

The first gene was and must be formed through selection of [GADV]-microshpere (the second motive force) having a higher growth rate, which was caused by more efficient synthesis of immature [GADV]-proteins (the first motive force) with a higher catalytic activity than others. Otherwise, the steps could not proceed towards the direction of formation of the first gene in random reactions on the primitive Earth. In other words, the two motive forces headed towards the direction of formation

of the first gene and resulted in the emergence of life, especially owing to the second motive force, [GADV]-microspheres (Figure 4).

(2) The first genetic system must be formed in parallel with formation of the first gene

The first reason is because both the first gene and the first genetic system were and must be formed through selection of [GADV]-microspheres, as described above. The second reason is because genetic information on the first gene never be expressed if the first gene was formed independently of formation of the first genetic system, and because the first gene becomes meaningless in the absence of the genetic system.

In addition, the genetic system never be formed if the first gene was formed independently of the genetic system, because it would be impossible to know the way how the genetic system or tRNA and genetic code, which mediate between gene and protein, can be formed. This means that it is impossible to form the first gene independently of the genetic system. The reverse would be also true.

(3) The reason why the first gene must be formed in a bottom-up manner

As shown in Figures 3 and 4, the first gene must be formed from immature [GADV]-protein (proto-mature protein) as piling up the rest of five proto-types of the members on the immature [GADV]-protein one by one. Formation of the four proto-members is indispensable to form the first gene in the bottom-up manner (Figure 6 (B)). The first genetic system, which are composed of four genuine members (gene, tRNA (genetic code) and protein), can use the respective proto-types of four members as genuine members, at the time when the first gene was formed (Figure 6 (B)). That is the reason why both the first gene and the first genetic system were formed in the bottom-up manner in parallel.

5. Discussion

Three matters described below are necessary to solve the origin of life. Note that, inversely stating this, it means that the “mystery” of origin of life can be solved, if the three matters could be reasonably explained.

(1) Comprehensive understanding of formation process of the fundamental life system.

(2) Understanding of formation process of the first gene.

(3) Understanding of formation process of the first genetic system.

Then, it is discussed below whether or not the three formation processes can be reasonably explained.

5.1. Comprehensive Understanding of Formation Process of the Fundamental Life System

Formation process of the fundamental life system composed of the six members must be not individually but comprehensively explained, because the six members are intimately related with each other. In GADV hypothesis, it is considered that formation processes of the six members advanced in a bottom-up manner or as starting from immature [GADV]-protein as shown in Figures 3, 4 and 6. Therefore, it can be understood that formation process of the fundamental life system can be reasonably explained as going upstream against the flow of gene expression (Figures 3, 4 and 6).

5.2. Understanding of Formation Process of the First Gene

The question, how the first gene was formed, has been one of the big problems in solving the origin of life for many years. On the contrary, in GADV hypothesis, it can be explained that the first gene was formed as progressing from immature protein to the first gene (Figures 3, 4 and 6).

5.3. Understanding of Formation Process of the First Genetic System

As a matter of course, formation of the first genetic system is also indispensable, because the genetic information encoded by the first gene never be expressed in the absence of the system. The process can be also reasonably explained in GADV hypothesis (Figure 6).

5.4. Why the First Genetic System Could Be Formed

As described in Subsection 4.2, it has been understood how the first gene was formed. However, life would not arise, even if only the first gene was acquired independently of the genetic system, because the genetic function could not be expressed in the microsphere, in which a genetic system did not exist. However, the first genetic system could be formed when the first gene was acquired in the bottom-up manner as described in Subsection 4.3. The reason is because the respective prototypes, which were formed for efficient synthesis of immature [GADV]-proteins in the absence of gene, could be used as genuine members of the first genetic system at the moment when the first gene could be formed (Figure 6 (B)).

All of the three formation processes of the fundamental life system, the first gene and the first genetic system can be explained reasonably. This means that the “mystery” of the origin of life could be solved by GADV hypothesis. In the following sections, it is discussed the reason why the “mystery” could be solved by GADV hypothesis.

5.5. Why Could the “Mystery” of the Origin of Life Be Solved by GADV Hypothesis

5.5.1. Three Keys for Solving the “Mystery” of Origin of Life

According to GADV hypothesis, it can be considered that the first life arose through the six steps described in Section 3 (Figures 3 and 4). At that time, three keys, immature [GADV]-protein, anticodon stem-loop RNA (proto-tRNA) and ss-(GNC)_n RNA (proto-mRNA), all of which were produced under random processes and played the respective important roles in the emergence of life (Figure 4, Table 1). That is, they were proto-types of protein (immature [GADV]-protein), tRNA (AntiC-SL RNA) and gene (ss-(GNC)_n RNA). The three keys contributed serially to form the fundamental life system, as the first key (immature [GADV]-protein) induced to form the second key (AntiC-SL RNA) and the second key triggered formation of the third key (ss-(GNC)_n RNA), as the former relayed a baton to the latter.

Table 1. Three keys for solving the mystery origin of life.

Three keys	Synthetic method	Roles of the key
1. imm-[GADV]-protein	random joining of [GADV]-amino acids	formation of cell membrane and proto-metabolism
2. AntiC-SL RNA	random joining-degradation cycles of nucleotides	formation of GNC code and ss-(GNC) _n RNA
3. ss-(GNC) _n RNA	random joining of GNCs in AntiC-SL RNAs	formation of ds-(GNC) _n gene; emergence of life

5.5.2. Two Motive Forces Heading to the Emergence of Life

Consider here the reason why the three keys, which generated the first gene, could be formed on the primitive Earth, on which only random reactions occurred, because generally speaking it would be impossible to form the first gene under random processes on the primitive Earth. On the other hand, only the processes, how the three keys were created, are explained in Section 3 (Figures 3 ad 4) without explanation of the reason, why the three keys could be created.

Then, the reason, why the the first gene could be formed, is explained here. The reason is owing to two motive forces, immature [GADV]-proteins and [GADV]-microspheres enclosing the immature proteins. The two motive forces, made it possible to proceed in the direction to the emergence of life (Figure 4, Table 2). Internal osmotic pressure of [GADV]-microspheres, which had a way synthesizing immature [GADV]-proteins with a higher catalytic activity more efficiently, increased at a higher rate than others. The higher internal osmotic pressure made it possible for the microspheres to grow and proliferate faster than others. Selection of such [GADV]-microspheres with a higher proliferation rate headed to the direction to the emergence of life, because, for example, formation of the three keys, immature [GADV]-protein (proto-protein), AntiC-SL RNA (proto-tRNA) and ds-(GNC)_n RNA

(proto-mRNA) gave a higher proliferation ability to [GADV]-microsphere. Consequently, the first gene could be formed in the selected [GADV]-microsphere. The two motive forces were founded on synthesis of immature [GADV]-proteins as aggregates of [GADV]-peptides or pseudo-replication of immature [GADV]-proteins in [GADV]-microsphere.

Table 2. Two motive forces, which headed to the direction of the emergence of life.

Two motive forces	Role of the motive forces
1. imm-[GADV]-protein	[GADV]-peptide synthesis increased osmotic pressure in [GADV]-microsphere
2. [GADV]-microsphere	selection of microspheres with a high proliferation ability headed towards the emergence of life

5.6. Comparison Between RNA World Hypothesis or GADV Hypothesis

Which hypothesis is valid as an idea for solving the “mystery” of the origin of life, RNA world hypothesis? or GADV hypothesis? As can be seen in Table 3, it would be difficult to synthesize nucleotides and RNAs on the primitive Earth because of their complex structures. On the contrary, it would be easy to produce [GADV]-amino acids and [GADV]-proteins on the primitive Earth, because the structures of [GADV]-amino acids and [GADV]-proteins are relatively simple.

Furthermore, self-replication of RNA is quite difficult and would be actually impossible. However, it would be possible to pseudo-replicate [GADV]-proteins owing to protein 0th-order structure. It would be also impossible to write genetic information for synthesis of a mature protein into RNA in the absence of protein, even if RNA could be self-replicated. On the contrary, it would be possible to acquire genetic information under protein-early theory or GADV hypothesis (Figures 3–6). In addition, it would be impossible to establish the first genetic system in a top-down manner, as expected by RNA world hypothesis. On the other hand, it would be possible to form the first genetic system in a bottom-up manner or GADV hypothesis as shown in Figure 6 (B). Taking it into consideration that formation processes of all of the five items shown in Table 3 are possible in the case of GADV hypothesis, it is concluded that the steps from chemical evolution to the emergence of life must progress as expected by GADV hypothesis (Figures 2–4 and 6).

Table 3. Comparison of RNA world hypothesis with GADV hypothesis. In this table, five items, monomer, polymer, replication, formation of gene or protein and formation of the genetic system, are compared between RNA world hypothesis and GADV hypothesis.

	RNA world hypothesis	GADV hypothesis
1. Monomer	Nucleotides (complex)	[GADV]-amino acids (simple)
2. Polymer	RNA (complex)	immature [GADV]-protein (simple)
3. Replication	self-replication (difficult)	pseudo-replication (possible)
4. Formation of gene or protein	gene-early theory (impossible)	protein-early theory (possible)

5.7. Three Possibilities of Formation Process of the First Genetic System

Three different ways are logically considered to explain how the first genetic system was formed, because the system is composed of three (or four) members, gene, tRNA (genetic code) and protein and, therefore, the formation must start from one of the three (or four) members (Figure 7).

Of course, it would be also impossible to form the first genetic system in the third manner assuming that the first genetic system was formed from tRNA (genetic code), because any mediator cannot be generated in the absence of objectives, which a mediator mediates (Figure 7). Therefore, the genetic system must be formed in the bottom-up manner according to the GADV hypothesis,

because the origin of life never be elucidated by RNA world hypothesis as shown in Table 3. In other words, these mean that the “mystery” of the origin of life has solved by the GADV hypothesis, because the genetic system must be naturally starts from either one of the three members composing the system. However, the genetic system cannot be formed from both gene located at the most upstream of the genetic system and tRNA (genetic code) located in the middle. This means that the formation must start from the remaining member, protein, located at the most downstream of the genetic system. That is immature but water-soluble globular [GADV]-proteins composed of simple and the most appropriate combination of [GADV]-amino acids for [GADV]-protein synthesis.

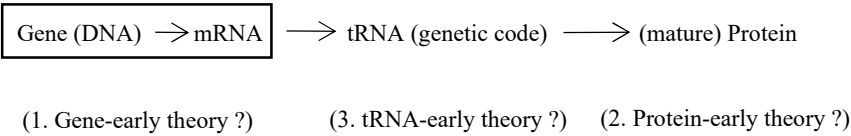


Figure 7. Three different ways for formation of the first genetic system can be logically considered. RNA world hypothesis and GADV hypothesis are grounded on the gene-early theory and protein-early theory, respectively. The third way or tRNA-early theory is impossible for mediator, tRNA, to bridge over between gene and protein in the absence of gene and protein.

In fact, the steps from immature [GADV]-proteins to the emergence of life can be reasonably explained (Figures 3 and 4). From the above considerations, it can be concluded that GADV hypothesis is the most valid idea for explaining the steps to the emergence of life.

Conclusion

RNA world hypothesis is one of gene/replicator-early theories. However, gene would not be first formed, because gene is a region of DNA or RNA, into which amino acid sequence information for synthesis of a mature protein is written. That is, gene itself cannot exhibit any concrete action. Therefore, life never emerge from RNA or gene, because nothing happen, even if such genes were first formed. Of course, I know that it is considered in RNA world hypothesis that life emerged from RNA world, which was formed by self-replication of ribozymes having a catalytic activity.

However, modern organisms are living by using proteins, which are working polymers. Therefore, the respective catalytic activities on RNAs must be transferred to the corresponding proteins one day, if the first life did really emerge from RNA world, and genetic information for a mature protein synthesis must be written into RNA strands, which were produced by self-replication. However, both transfer of catalytic activity on RNA strand onto protein with three-dimensional structure and writing genetic information into RNA strand are principally impossible. This means that, even if RNA could be self-replicated, the self-replicated RNAs must remain in the RNA world and, in other words, that RNA world cannot evolve into RNA-protein world.

Contrary to that, it can be considered that life emerged from [GADV]-protein world, which was formed by pseudo-replication of immature [GADV]-proteins having catalytic functions. Stating this more concretely, it is considered that the first life arose as one of [GADV]-microspheres holding immature [GADV]-proteins and that the first genetic system was established at the moment, when the first (GNC)_n gene was formed through selection of microspheres having more efficiently synthesizing system of immature [GADV]-proteins with higher catalytic activities. Therefore, it can be concluded that the “mystery” has been solved by GADV hypothesis [12,13], considering both formation of immature [GADV]-proteins (the first motive force) under one of protein 0th-order structure or [GADV]-amino acids and selection of [GADV]-microspheres (the second motive force) proliferating with a higher rate than others.

Funding: This research received no external funding.
Informed Consent Statement: Not applicable.
Conflicts of Interest: The authors declare no conflict of interest.

Acknowledgments: 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 origin and evolution of the fundamental life system.

References

1. Gilbert, W.; The RNA world. *Nature*, **1986**, 319, 618.
2. Saito H. The RNA world 'hypothesis'. *Nat Rev Mol Cell Biol.* 2022;23:582.
3. Fine JL, Pearlman RE. On the origin of life: an RNA-focused synthesis and narrative. *RNA*. 2023;29:1085–1098.
4. Corliss JB., Dymond J, Gordon LI, Edmond JM, von Herzen RP, Ballard RD, Green K, Williams D, Bainbridge A, Crane K, van Andel TH. Submarine thermal springs on the galapagos rift. *Science*. 1979;203:1073–1083.
5. Holm NG, Andersson E, Hydrothermal simulation experiments as a tool for studies of the origin of life on Earth and other terrestrial planets. A review. *Astrobiology*, 2005;5:444–460.
6. Zhang X, Tian G, Gao J, Han M, Su R, Wang Y, Feng S. Prebiotic synthesis of glycine from ethanolamine in simulated Archean alkaline hydrothermal vents. *Orig Life Evol Biosph.* 2017;47:413–425.
7. Arrhenius S. Evolution of the universe. 1908; Harper, London, United Kingdom.
8. Temple R. The prehistory of panspermia: astrophysical or metaphysical? 2007; Cambridge University Press, Cambridge, United Kingdom.
9. Furukawa Y, Chikaraishi Y, Ohkouchi N, Ohkouchi N, Ogawa NO, Glavin DP, Dworkin JP, Abe C, Nakamura T. Extraterrestrial ribose and other sugars in primitive meteorites. *Proc. Natl. Acad. Sci.* 2019;116:24440–24445.
10. Ikehara, K. Protein ordered sequences are formed by random joining of amino acids in protein 0th-order structure, followed by evolutionary process. *Orig. Life Evol. Biosph.* **2014**, 44, 279–281.
11. Ikehara, K. The first genetic system was established not in the top-down manner (RNA world hypothesis) but in the bottom-up manner (GADV hypothesis). *Medical Research Archives*, **2024**, 12, doi.org/10.18103/mra.v12i7.5457
12. 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.
13. Ikehara, K.; Possible steps to the emergence of life: The [GADV]-protein world hypothesis. *Chem. Rec.*, **2005**, 5, 107–118.
14. Ikehara, K.; Okazawa, E. Unusually Biased Nucleotide Sequences on Sense Strands of *Flavobacterium* sp. Genes Produce Nonstop Frames on the Corresponding Antisense Strands. *Nucleic Acids Res.*, **1993**, 21, 2193–2199.
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**, 24:4249–4255.
16. Ikehara, K.; Yoshida, S. SNS hypothesis on the origin of the genetic code. *Viva Origino* **1996**, 26, 301–310.
17. Ikehara, K.; Omori, Y.; Arai, R.; Hirose, A. A novel theory on the origin of the genetic code: a GNC-SNS hypothesis. *J. Mol. Evol.* **2002**, 54, 530–538.
18. Ikehara. K. Pseudo-replication of [GADV]-proteins and Origin of Life, *Int. J. Mol. Sci.*, **2009**, 10, 1525–1537.
19. Ikehara, K. *Towards Revealing the Origin of life.—Presenting the GADV Hypothesis*; Springer Nature, Gewerbestrasse: Cham, Switzerland, **2021**.
20. Ikehara, K. The origin of tRNA deduced from *Pseudomonas aeruginosa* 5' anticodon-stem sequence: Anticodon stemloop hypothesis. *Orig. Life Evol. Biosph.* **2019**, 49, 61–75.

21. Shimizu, M. Molecular basis for the genetic code. *J. Mol. Evol.* **1982**, 18, 297-303.
22. Yarus, M. The genetic code and RNA-amino acid affinities. *Life* **2017**, 7, 13.
23. Ikehara, K. Why were [GADV]-amino acids and GNC codons selected and how was GNC primeval genetic code established? *Genes* **2023**, 14, 375.
24. Van der Gulik, P.; Massar, S.; Gilis, D.; Buhrman, H.; Rooman, M. The first peptides: the evolutionary transition between prebiotic amino acids and early proteins. *J. Theor. Biol.*, **2009**, 261, 531-539.
25. 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>

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.