Preprints.org

Review

Why Could Life Emerge? -Gadv Hypothesis

Kenji Ikehara *

Review

Why Could Life Emerge? -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 "Mystery" of the origin of life has not been solved still now. In this review article, the reason, why the origin of life has not been solved, is discussed from three standpoints, (1) formation process of the first gene, (2) formation process of the first genetic system, (3) formation process of the first fundamental life system. From the results, it has been understood that the main reasons are because the studies on the origin of life have been chiefly carried out under the RNA world hypothesis based on the "gene/replicator-first hypothesis". Successively, it was investigated whether or not the "mystery" of the origin of life could be solved by [GADV]-protein world hypothesis or GADV hypothesis based on the "protein/metabolism-first hypothesis". Consequently, it has been confirmed that the steps to from chemical evolution to the emergence of life can be reasonably explained, if the formation process of the first gene was started from not mature but immature [GADV]-protein, actually [GADV]-peptide aggregates. It was also found that the three problems (1)~(3) is one problem, that is, the second and the third problems can be simultaneously solved, if the first problem (1), "How was the first gene formed?" could be solved. Furthermore, it has been found that the formation process of the first fundamental life system leading to the emergence of life must be formed by piling up the five members onto the immature [GADV]-protein according to the bottom-up manner.

Keywords: Origin of life; RNA world hypothesis; [GADV]-protein world hypothesis; GADV hypothesis; Origin of gene; Origin of genetic system; Origin of fundamental life system

1. Introduction

Human beings have been interested in the "Mystery" of the origin of life about the process, through which the first life arose, for thousands of years. The problem about the origin of life, "How did life emerge on the primitive Earth?", had been long disputed until about 20 years ago. One of the representatives is the dispute, which was discussed between "gene/replicator-first theory (in short, gene-first theory)", under which many researchers took a firm stand for RNA world hypothesis [1-3], and "protein/metabolism-first theory (in short, protein-first theory)", under which a few researchers was actively opposed the RNA world hypothesis [4-7]. For example, Shapiro had argued against the RNA world hypothesis from the standpoint of protein-first theory [8-10]. However, the disputes eventually converged and the origin of life have been mainly studied under the RNA world hypothesis up to the present. However, the "mystery" has not been solved still now. There are also other disputes about the fields of the origin of life, such as "Hydrothermal vent theory" advocating that life emerged around volcano in a deep sea [11-13]. However, the "mystery" of the origin of life could not been solved, because the mechanism, how the first genetic system was established, could not be explained by the idea. Therefore, this review article begins from descriptions of the reasons, why the "mystery" of the origin of life has not been solved still now.

Then, it is considered first what are the matters indispensable for modern organisms to live, because, by doing so, the "mystery" of the origin of life could be solved by revealing formation processes of the matters, which should be indispensable for the first life to arise on the primitive Earth.

2. What Matters Are Indispensable for Modern Organisms To Live?

Many persons would agree with that both "gene" and "protein" are indispensable for modern organisms to live (Figure 1). "Genetic code", which mediates between gene and protein, and "tRNA" realizing the relationship between codon and amino acid displayed by the genetic code, must be indispensable (Figure 1). Therefore, the genetic system composed of gene, genetic code (tRNA) and protein would be indispensable. Many researchers have tried to make clear how the genetic system was formed. However, many studies on formation of the genetic system have been carried out not comprehensively but rather individually regardless of the system, which is in a united body. Consequently, formation process of both the respective members and the genetic system itself have not been made clear still now.

On the other hand, the genetic system itself is, of course, not life, because the genetic system put into a test tube cannot be called as "life". Life is a matter, which is surrounded by "membrane" and can divide and proliferate. "Metabolic system" is also indispensable for modern organisms to live, because increase of internal osmotic pressure, which is induced by polymer syntheses, is necessary for cells to proliferate (Figure 1). Therefore, it would be valid to consider that modern organisms are living under the fundamental life system, which is composed of six members, gene, genetic code, tRNA, metabolism, cell structure and protein.

As described above, it would be necessary to make clear formation processes of the first gene, the first genetic system and the first fundamental life system in order to solve the "mystery" of the origin of life, although the three are deeply interrelated. Note that the reason, why the three on different levels are enumerated above, is because the "mystery" of the origin of life must not be revealed, if even the origin of the simplest one, gene, could not be made clear.

3. Three Minimum Requirements for Solving the "Mystery" of the Origin of Life

As described above (Section 2), solving the "mystery" of the origin of life is to make clear the formation processes of the first matters, which are indispensable for life to live. Then, enumerate again the three first matters necessary to solve the "mystery" of the origin of life.

3.1. Formation Process of the First Gene (Figure 1 (1)).

3.2. Formation Process of the First Genetic System Composed of Gene, Genetic Code (tRNA) and Protein (Figure 1 (2)).

3.3. Formation Process of the First Fundamental Life System Composed of Gene, Genetic Code, tRNA, Metabolism, Cell Structure and Protein (Figure 1 (3)).

The reason, why the formation process of the first gene but not the first protein is raised as the question above, is because many researchers are studying on the origin of life under the gene-first theory but not the protein-first theory, and also because at this time point it is considered that the "mystery" of the origin of life could be elucidated by RNA world hypothesis. As a matter of course, gene is here a mature gene encoding an amino acid sequence of a mature or refined protein. The reason, why a mature gene is emphasized here, is because it should be clearly differentiated a mature gene from an immature gene encoding an immature or unrefined protein, which can be produced by random joining of [GADV]-amino acids in the absence of gene and also because the immature genes and immature proteins, which played important roles in the emergence of life, are frequently discussed in the later Sections.
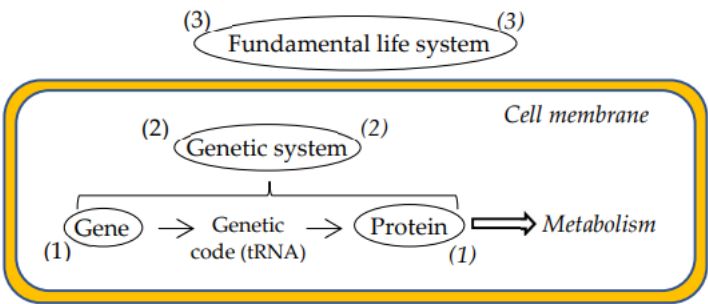
3 of 15

Figure 1. Three matters necessary to solve the "mystery" of the origin of life; Gene (1), Genetic system (2) and Fundamental life system (3) under the RNA world hypothesis or in top-down manner [13]. In the case of GADV protein world hypothesis (GADV hypothesis) or bottom-up manner, the three matters are changing from Gene (1) to Protein *(1)* as shown in italic numbers.

Next, it is discussed whether or not formation processes of the three matters (1. gene, 2. genetic system and 3. fundamental life system) can be reasonably explained based on the "gene-first theory" or RNA world hypothesis one by one in order to understand the reason, "why the "mystery" of the origin of life has not been solved still now?".

4. Can the Three Formation Processes be Explained by RNA World Hypothesis or in a Top-Down Manner?

4.1. Can Formation Process of the First Gene be Explained in the Absence of Genetic Code (tRNA) and Protein?

As well known, the most important matters for modern organisms to live are undoubtedly "(mature) gene" and "(mature) protein" composing of the genetic system. Gene is a region on a DNA or RNA strand, into which an amino acid sequence of a refined protein is written. However, even the problem, "how was gene formed?", in other words, "how was genetic information for protein synthesis written into DNA (or RNA) as codon sequence?", could not be uncovered still now. Or rather, it would be impossible in principle to write any amino acid sequence of a mature protein into DNA (RNA) in the absence of protein, because it is logically impossible for the DNA (RNA) to know even a protein, of which amino acid sequence should be written, in the absence of protein. Thus, it can be confirmed that a mature gene never be formed in the absence of protein. The reason, why it is daringly reconfirmed here that gene never be formed in the absence of protein, is because RNA world hypothesis, under which many researchers are studying on the origin of life, is based on the gene-first theory and moreover the hypothesis is a mainstream of the studies on the origin of life still now.

4.2. Can the First Genetic System Be Formed?

Next, consider how the first genetic system was formed. There are naturally three possibilities in the formation process of the first genetic system, because the system was composed of gene, genetic code (tRNA) and protein (Figure 2).

4.2.1. The First Possibility: Could the First Genetic System Be Formed from Gene? (Figure 2 (1))

As explained in Introduction, RNA world hypothesis is an idea based on the "gene-first theory". Therefore, needless to state, it is considered in the hypothesis that gene was formed before formation of protein and establishment of the genetic code (tRNA) and life emerged after establishment of the first genetic system. However, as described above (Subsection 4.1), gene having information for a mature protein synthesis never be formed. Nevertheless, it seems to me that many persons are implicitly considering that the establishment process began from gene. The reason would be because proteins are always synthesized through gene expression in modern organisms. Surely, protein never be produced without gene as long as the protein is a mature or refined protein.
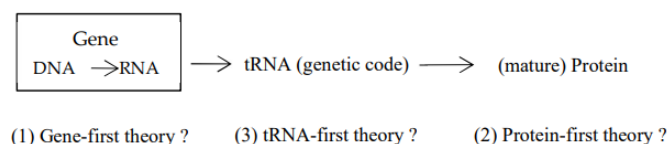
Figure 2. Three formation processes of the first genetic system can be logically considered. The first one is an idea that the formation process began from gene, as considered in RNA world hypothesis and the second one is an idea that the formation process began from protein, as considered in GADV hypothesis. The two hypotheses are grounded on the gene-first theory and protein-first theory, respectively. The third one is, so to state, "tRNA-first theory", an idea that formation of the genetic system started from a mediator, tRNA, which bridges over between gene and protein, in the absence of gene and protein.

As well known, RNA world hypothesis was proposed by W. Gilbert about 40 years ago to solve the difficult question; the "chicken-egg relationship" between gene and protein [1]. The discovery of ribozymes or catalytic RNAs triggered the proposition of RNA world hypothesis [14,15], because it was considered that formation process of the genetic system could be explained by transferring genetic information and catalytic activity on a self-replicated RNA to DNA and protein, respectively. Note that it has been widely considered that the genetic system is composed of gene-mRNA-protein according to the so-called "Central dogma", which was presented by F. Crick [16]. Therefore, the genetic code and tRNA are not explicitly appeared in the "Central dogma", in which (m)RNA is overestimated. For the reason, I have proposed the "Core (genetic) system" (Figure 2) in order to modify the overestimate of mRNA and the underestimate of tRNA (genetic code) in the Central dogma.

Thus, the RNA world hypothesis is an idea considering that the first genetic system was formed along with the flow of gene expression (in a top-down manner). The idea is based on the fact that genetic information written into DNA or RNA is expressed with tRNA (genetic code). That is, the idea matches the fact that (mature) protein never be synthesized in the absence of gene. Therefore, it would be unavoidable for many persons to consider intuitively that the genetic system must be formed along with the flow of gene expression or from gene. However, if it is well considered, it should be easily understood that it is impossible to form both the first gene and, therefore, the first genetic system in the top-down manner as described above (Subsection 4.1).

Nonetheless, the studies on the origin of life have been carried out under RNA world hypothesis, which is based on the "gene-first theory" still now. It seems to me that, therefore, the studies on the origin of life have been carried out with a focus on construction of self-replicated RNA but not on formation of the first gene [17], although even such a self-replicated RNA has not been constructed in a real sense still now. It would be impossible to solve the "mystery" of the origin of life as long as the studies on the origin of life are carried out from the standpoint of RNA world hypothesis or the gene-first theory.

4.2.2. The Second Possibility: From Protein

Then, consider the second possibility, whether or not the first genetic system can be formed and the "mystery" of the origin of life can be solved, if formation process of the genetic system was begun from (mature) protein in the absence of gene and genetic code (tRNA) (Figure 2 (2)). However, it is unquestionably impossible to form a mature protein in the absence of gene, because any mature protein must be synthesized under genetic function.

4.2.3. The Third Possibility: From tRNA (Genetic Code)

Discuss here whether or not the first genetic system could be formed from tRNA (genetic code) (Figure 2 (3)) It would be also logically impossible to form tRNA in the absence of gene and protein. The reason is because it is impossible to form tRNA mediating between gene and protein in the absence of the two objects, gene and protein. Nevertheless, the reason, why it is discussed about the possibility in this article, whether or not the genetic system can be formed from tRNA (genetic code), because the genetic system must be formed from formation of one of the three members (Figure 2).

Next, consider whether or not tRNA can be formed, if self-replicated RNA could be produced. Only meaningless RNA, but not tRNA, might be formed independently of gene and protein. However, it would be also impossible for an imaginary proto-tRNA to know even the way how to transmit genetic information for protein synthesis to amino acid sequence in the absence of gene and protein. Therefore, there must be no way for generating a meaningful proto-tRNA.

The reason, why the sentences above have been daringly described here too, is because many researchers might be studying on the origin of life under an idea that it would be possible to form the first genetic system from not only gene and protein but also tRNA (genetic code).

4.3. Can "The Fundamental Life System" Be Formed?

Of course, it would be impossible to reveal formation process of the fundamental life system containing the genetic system, because the origin of the genetic system composed of gene, genetic code (tRNA) and protein cannot be made clear as described Subsection 4.2. In such situation, an article entitled as "*The RNA world hypothesis: the worst theory of the early evolution of life (except for all the others)*" was published by Harold S. Bernhardt in 2012 [18]. It is described in the article that "*any idea beyond RNA world hypothesis had not been presented, although RNA world hypothesis has a huge problem*". Contrary to that, it has been repeatedly explained in sections above that formation process of the genetic system cannot be explained by RNA world hypothesis based on the gene-first theory. On the other hand, there is the fact that life emerged about 4 billion years ago, as it is easily understood from that splendid organisms are living on the present Earth. The fact clearly shows that life emerged after the first gene, the first genetic system and the first fundamental life system were formed in some way.

5. Steps to the Emergence of Life Viewed from GADV Hypothesis

How did life emerge on the primitive Earth? For solving the "mystery", it is necessary to understand the two concepts, the first one is protein 0th-order structure [19] or [GADV]-amino acid composition and the second one is pseudo-replication of [GADV]-protein [20]. The purpose of this review article is to solve the "mystery" of the origin of life from the standpoint of GADV protein world hypothesis or GADV hypothesis [21-23], which I have proposed based on the protein 0th-order structure and pseudo-replication of [GADV]-protein.

Then, consider whether or not steps to the emergence of life, or the three formation processes of the first protein, the first genetic system and the first fundamental life system, can be reasonably explained, if the processes are viewed from GADV hypothesis.

The studies, which led to the GADV hypothesis, began from investigation on formation of entirely new genes in modern organisms about 25 years ago [22,23], when the RNA world hypothesis was in a golden age. Needless to state, GADV hypothesis is one of protein world hypotheses, which eventually lost an argument with the RNA world hypothesis. Therefore, many persons may consider that it must be impossible to explain the steps to the emergence of life from standpoint of the protein-first theory, because the problem has already settled. However, it should be noted that the argument was settled but without taking GADV hypothesis into consideration.

It is examined whether or not the three formation processes can be explained according to the GADV hypothesis, similarly as the case considered in the RNA world hypothesis (Section 4), although, of course, the "formation process of the first protein" must be considered under the GADV hypothesis instead of the "the first gene" under the RNA world hypothesis.

(5.1) Formation process of the first protein (immature or unrefined protein) (Figures 1 and 2).

(5.2) Formation process of the first genetic system composed of protein, genetic code (tRNA) and gene (Figures 1 and 2)..

(5.3) Formation process of the first fundamental life system composed of protein, cell structure, metabolism, tRNA, genetic code and gene (Figure 1).

5.1. Formation Process of the First Protein (Immature or Unrefined Protein)

Surely, GADV hypothesis is an idea based on the "protein-first theory" [21-23]. Of course, it is obvious that formation process of protein never be explained even from the standpoint of GADV hypothesis, if the protein is a mature or refined protein, which is synthesized under genetic

expression system. It is impossible to produce such a mature protein in the absence of gene because of a high wall of 10^{130}, which it is impossible to overcome [24]. On the other hand, as describe in Subsection 4. 1, gene never be formed in the absence of protein, because both synthesis of gene and expression of genetic information requires the presence of protein. This is the "chicken-egg" relationship observed between gene and protein.

This means that there is a large misunderstanding in some way. Many persons have overlooked an important concept, protein 0th-order structure, which makes it possible to form immature but water-soluble globular proteins with a large flexibility in the absence of gene or the genetic system [19]. Therefore, it is considered in the GADV hypothesis that the first proteins were immature [GADV]-proteins, actually [GADV]-peptide aggregates, which were formed by direct random joining of [GADV]-amino acids in the absence of gene. In addition, it is important to know that formation of the immature [GADV]-proteins triggered formation of the first ds-$(GNC)_n$ RNA encoding an immature [GADV]-protein and further eventually led to the formation of a mature $(GNC)_n$ gene encoding a mature [GADV]-protein for understanding steps to the emergence of life, as explained in the following Subsections in detail.

5.2. Formation Process of the First Genetic System

As already explained in Section 4.2, it is impossible to form the genetic system even if formation process of the genetic system began from gene or tRNA (genetic code). However, protein is also contained in the genetic system. This means that only one possibility, or beginning formation of the genetic system from protein, is left, although, of course, it is unknown whether or not it is really possible to explain the formation process of the genetic system, if the process started from protein. Therefore, it is the purpose to confirm whether formation process of the genetic system can be reasonably explained from standpoint of the "protein-first theory" or not.

For the purpose, it is important to make clear origin of the genetic code, because it can be understood what were the first gene and the first protein, if it could be known what was the first genetic code. For example, it can be considered that the first gene and the first protein were $(GNC)_n$ gene and [GADV]-protein, respectively, if the universal genetic code originated from GNC primeval genetic code (Figure 3 (A)) [25,26]. Then, first confirm whether or not the GNC primeval genetic code hypothesis is a valid idea for the origin of the genetic code.

Many researchers have recently agreed with the idea that the genetic code originated from GNC primeval genetic code. As the reason, some evidences showing that the first genetic code was GNC code, can be provided as follows.

[GADV]-amino acids can satisfy the four conditions (hydrophobicity/hydrophilicity, α-helix, β-sheet and turn/coil formabilities) for formation of water-soluble globular protein, which were obtained with amino acid compositions [25]. Note that four amino acids extracted from each row and each column of the universal genetic code table did not satisfy the four conditions, except [GAEV]-amino acids, which are similar to the [GADV]-amino acids [22]. [E] means glutamic acid.

[GADV]-amino acid composition is, therefore, one of protein 0th-order structure. This means that not only the genetic code simply represents correspondence relationship between codons and amino acids, but also protein 0th-order structure, which is necessary to produce water-soluble globular proteins under random joining of amino acids, is written into the genetic code table. Owing to the protein 0th-order structure, [GADV]-amino acids could be selected from organic compounds containing various amino acids, which were synthesized with prebiotic means [26].

On the other hand, GNC codons were selected from various combinations of nucleotides. The reason is because triplet base pairs, which were formed between two complementary GNCs turned out from the loop of anticodon stem-loop (AntiC-SL) RNA, or GGC/GCC and GAC/GUC are more stable than other triplet base pairs [27]. Such properties made it possible to use GNC as anticodon carried by tRNA [26].

Correspondence relationships between GNC codons and [GADV]-amino acids were determined by accidentally freezing within the limits of two double pairs of GGC/GCC, GAC/GUC and Gly/Ala, Asp/Val [21,26].

Thus, it can be confirmed that the first genetic code was GNC code connecting GNC codons with [GADV]-amino acids. It can be anticipated that the first genetic information was written into ds-RNA as ds-$(GNC)_n$ RNA encoding the first protein, which was composed of [GADV]-amino acids.

Then, next consider whether or not formation processes of the genetic system can be reasonably explained beginning from immature [GADV]-protein, if the concept, protein 0th-order structure or [GADV]-amino acid composition, was taken into considerations.
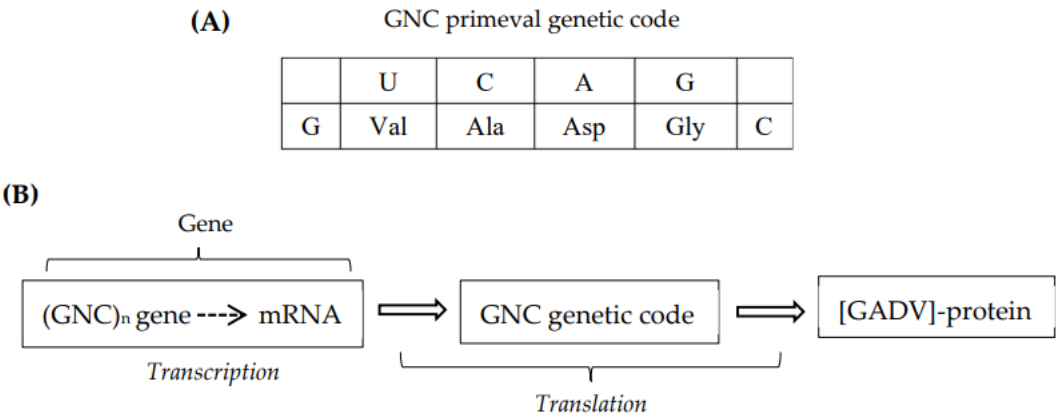

Figure 3. (A) GNC primeval genetic code. The triplet code encoding four [GADV]-amino acids. **(B)** The genetic system composed of $(GNC)_n$ gene, GNC genetic code and [GADV]-protein. It is considered that the "mystery" of the origin of life could be solved by revealing formation process of the genetic system.

5.3. Formation Process of the First Genetic System Deduced from GADV Hypothesis

As describe above, the first proteins were immature or unrefined [GADV]-proteins, which were actually [GADV]-peptide aggregates. Then, explain formation process of the first genetic system according to the bottom-up manner (Figure 4). Only the main points, which were obtained in the early studies on the origin of life [22,23], are drawn in the Figure.

It is deduced that the first proteins were composed of [GADV]-amino acids based on the GNC primeval genetic code (Figure 3 (A)).

It is considered that the first [GADV]-proteins could be produced in the absence of gene, because [GADV]-amino acids having relatively simple chemical structures could be easily synthesized with prebiotic means and should accumulate at large amounts on the primitive Earth.

[GADV]-polypeptide chains could be folded into unrefined but water-soluble globular structures with a large flexibility, because the immature protein can satisfy the four conditions for protein structure formation.

The immature [GADV]-proteins could be used as proteinaceous catalysts with weak but diverse activities because of a large flexibility.

Therefore, [GADV]-amino acids and nucleotides could be synthesized by the immature [GADV]-proteins, which were produced by pseudo-replication of immature [GADV]-proteins [20].

Finally, the first life emerged after establishment of GNC primeval genetic code and formation of the first $(GNC)_n$ genes.

When a scenario of steps to the emergence of life was obtained for the first time, I only just noticed one type of protein world hypotheses as drawn in Figure 4 [22,23]. The scenario is that the first life emerged beginning from pseudo-replication of immature [GADV]-proteins via establishment of GNC primeval genetic code and formation of $(GNC)_n$ primeval gene. Although only the outline from chemical evolution to the emergence of life could be obtained at that time, it has been now confirmed that the scenario is consistent with the present ideas about the emergence of life, which are shown in the Figures 5 and 6 [21].
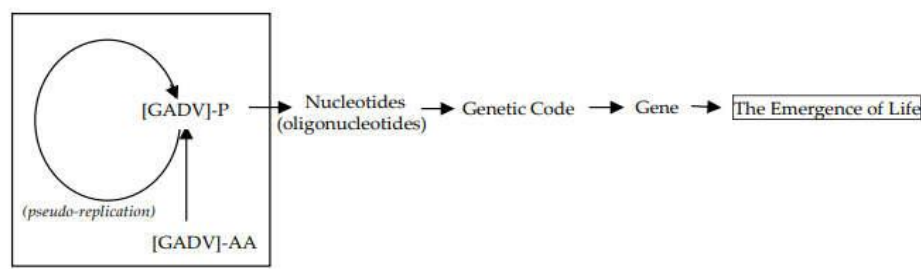
Figure 4. Steps from pseudo-replication of [GADV]-protein to the emergence of life, which was assumed when GADV hypothesis was hit upon by accident.

5.4. Formation Process of the Fundamental Life System

As can be seen below, steps from chemical evolution to the emergence of life can be explained by GADV hypothesis. The reason, why life arose starting from immature [GADV]-proteins, which were produced by random joining of [GADV]-amino acids in the absence of gene and finally arriving at formation of the first gene as piling up the five members onto the immature [GADV]-proteins, can be also explained (Figure 5).
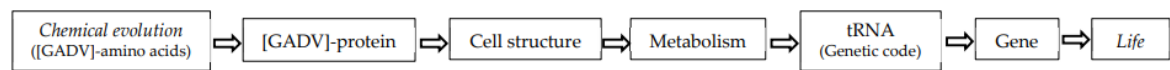

Figure 5. Steps from chemical evolution to the emergence of life, which was considered based on recent studies on the origin of life. Formation processes of the six members, protein, cell structure, metabolites, tRNA, genetic code and gene, can be comprehensively explained.

I have now confidence in the GADV hypothesis suggesting that life did emerge through the processes progressing against the flow of gene expression or in a bottom-up manner. It is discussed in the next Section in more detail, why the genetic system must be formed in the bottom-up manner and how the first life emerged.

6. How Did Life Emerge on the Primitive Earth?

The GADV hypothesis suggests that the first genetic system was formed by piling up five members onto immature [GADV]-proteins (Figure 5) [21]. In this Section, steps to the emergence of life, which are considered now in the GADV hypothesis, are explained (Figure 6).

6.1. Steps from Chemical Evolution to the Emergence of Life

Chemical evolution

During chemical evolution, [GADV]-amino acids were synthesized with prebiotic means. Miracle 1: [GADV]-amino acids, which could be synthesized with prebiotic means, had existed on the primitive Earth.

Immature [GADV]-protein

Immature [GADV]-protein, actually [GADV]-peptide aggregates, could be formed by direct random joining of [GADV]-amino acids, which accumulated on the primitive Earth. Miracle 2: Formation of such immature [GADV]-proteins became possible owing to the protein 0th-order structure or [GADV]-amino acid composition.

Formation of [GADV]-microsphere

[GADV]-microspheres were formed by association of the immature [GADV]-proteins. Miracle 3: [GADV]-microspheres could be formed with only immature [GADV]-proteins.

Formation of metabolic system

Primeval metabolic system was formed using catalytic activities of immature but pluripotent [GADV]-proteins in [GADV]-microspheres. In the metabolic system driven by immature [GADV]-proteins, [GADV]-amino acids could be produced in [GADV]-microsphere. Miracle 4: Primeval metabolic system could be formed with immature [GADV]-proteins in [GADV]-microspheres. Glyoxylate and pyruvate, which accumulated in surroundings of [GADV]-microspheres, could be used as precursor molecules for synthesis of [GADV]-peptide in the microspheres.

Nucleotide synthesis

Similarly, ribose 5-phosphate, ATP and other nucleotides were synthesized with immature [GADV]-proteins by using glyceraldehyde as a precursor molecule in [GADV] microspheres. Miracle 5: The trimer, $5'$-CCA$3'$, as an activator for [GADV]-peptide synthesis, also could be produced during cycles of random polymerization of nucleotides and degradation of oligonucleotides.
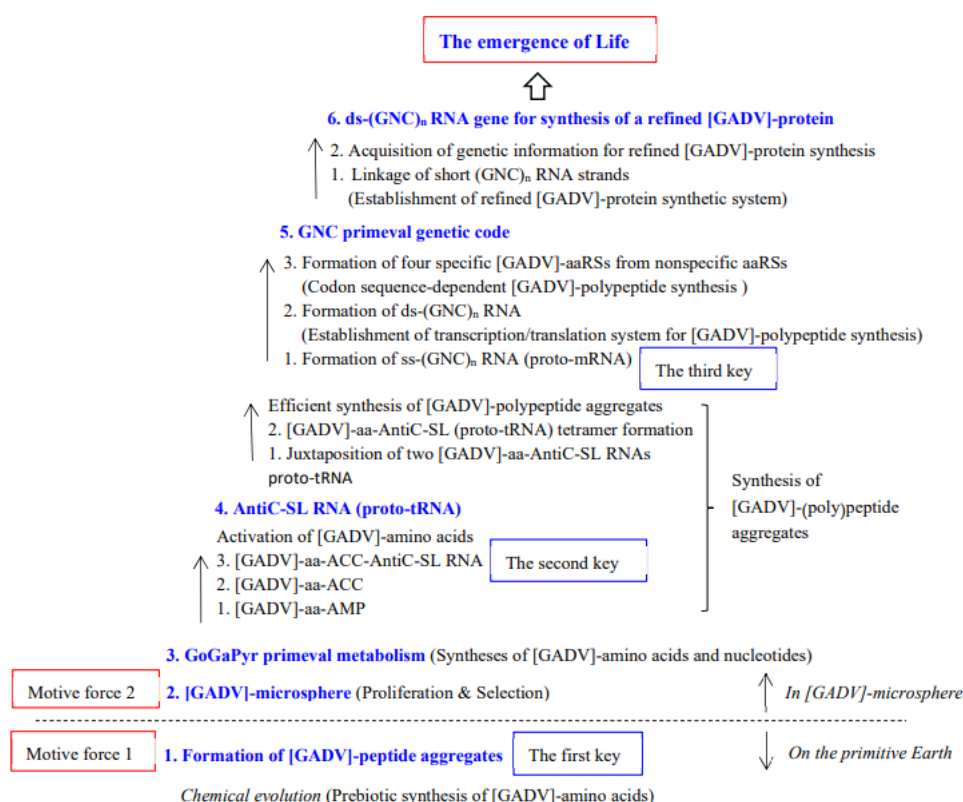

Figure 6. The steps from chemical evolution to the emergence of life. It can be understood that the first life arose owing to the three keys, immature [GADV]-proteins (actually [GADV]-peptide aggregates), AntiC-SL RNA and ss-(GNC)$_n$ RNA, and two motive forces, immature [GADV]-proteins (actually [GADV]-peptide aggregates) and [GADV]-microsphere.

Formation anticodon stem-loop (AntiC-SL) RNA (proto-AntiC-SL tRNA)

AntiC-SL RNAs were produced during repeated random joining of nucleotides and degradation cycles and [GADV]-peptide synthesis accelerated further by using AntiC-SL RNAs. Miracle 6: The AntiC-SL RNAs composed of 17 nucleotides could be formed during cycles of random polymerization of nucleotides and degradation of oligonucleotides. Miracle 7: In addition, random [GADV]-peptides could be synthesized by non-specific [GADV]-aminoacyl(aa)-AntiC-SL RNA. Furthermore, [GADV]-peptides could be synthesized by [GADV]-aa-AntiC-SL RNA dimers, which were formed by juxtaposition of two non-specific [GADV]-aa-AntiC-SL RNAs. Two [GADV]-aa-AntiC-SL RNA dimers could be, further, arranged vertically to make tetramer, by which [GADV]-peptides could be more efficiently synthesized than before.

Formation of ss-(GNC)$_n$ RNA and ds-(GNC)$_n$ RNA

ss-(GNC)$_n$ RNA was formed by random joining of GNCs carried by AntiC-SL RNA tetramers. Miracle 8: ss-(GNC)$_n$ RNA could be formed by random joining of anticodons, GNCs, which were

carried by AntiC-SL RNA tetramers. Arrangement of anticodons, GNCs, without gaps made it possible to form commaless ss-$(GNC)_n$ RNA sequence. After formation of ss-$(GNC)_n$ RNA, ds-$(GNC)_n$ RNA could be formed by complementary strand synthesis of ss-$(GNC)_n$ RNA.

Establishment of GNC primeval genetic code

GNC primeval genetic code was established, after correspondence relationship between [GADV]-amino acids and GNC anticodons, which were carried by AntiC-SL RNAs, was accidentally determined and frozen [21]. Miracle 9: The first GNC primeval genetic code could be established by accidentally freezing between four [GADV]-amino acids and four GNC codons.

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

The first ds-$(GNC)_n$ RNA gene encoding a refined [GADV]-protein was formed by maturation of a weak catalytic activity on an immature [GADV]-protein, which was produced by expression of either one of strand of ds-$(GNC)_n$ RNA [21]. Miracle 10: The first ds-$(GNC)_n$ RNA gene could be formed. In addition, both homologous genes and entirely new genes could be generated from $(GNC)_n$ codon sequences on sense strand and on antisense strand of the first ds-$(GNC)_n$ RNA gene, respectively.

The emergence of life

The first life emerged after a number of genes encoding a mature [GADV]-protein, which is necessary for the first life to live, were formed. Miracle 11: The fact, that the steps to the emergence of life had existed, is truly a miracle.

6.2. Eleven Miracles Made it Possible for Life to Emerge

It is considered that the first life arose on the primitive Earth through evolutionary process from chemical evolution to the emergence of life due to eleven miracles as described Subsection 6.1. Thus, the first life emerged after a number of genes necessary for the first life to live were acquired. The reason, why the word, "miracle", is used in the respective items described above, is because it seems to me as if the steps had previously prepared prior to the next steps, respectively. Therefore, it is not stated that phenomena, which are unlikely to occur, miraculously occurred.

6.3. Main Points of the GADV Hypothesis on the Emergence of Life

The steps to the emergence of life started from immature [GADV]-proteins (actually, [GADV]-peptide aggregates).

The first gene could be formed by piling up five proto-members (proto-cell structure, proto-metabolism, AntiC-SL RNA (proto-tRNA), proto-GNC code and proto-gene (ds-$(GNC)_n$ RNA)) onto the immature [GADV]-proteins, one by one, without perception of formation of the genetic system. That could be easily understood from that the piling-up of each member was carried out in the absence of gene. That is, the respective proto-members were used only for promoting efficient synthesis of immature [GADV]-proteins having a higher catalytic activity. That would be easily understood from that immature [GADV]-protein was always synthesized at the each step [28], as can be seen in Figure 7 (B)

The proto-members including ds-$(GNC)_n$ RNA changed into respective authentic members for gene expression at that moment, when the first ds-$(GNC)_n$ RNA gene was formed (Figure 7 (B). Of course, function of the respective members composing the first   genetic system was quite low at the time point, when the system was completed. However, the respective members were refined to ones having higher functionality through selection of [GADV]-microspheres with proliferation ability higher than others.
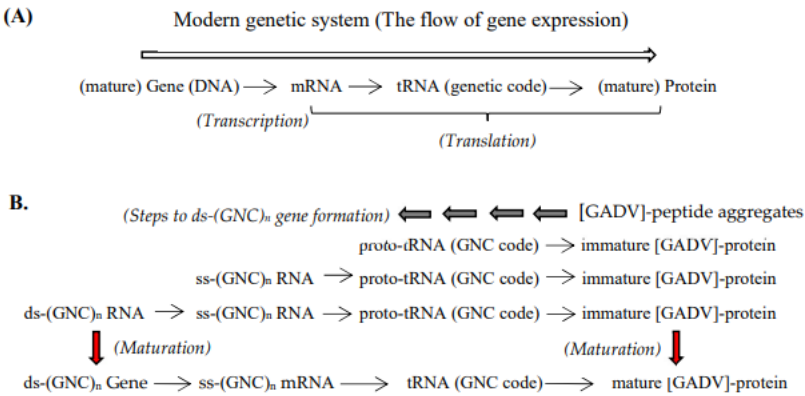
Figure 7. (A) Modern genetic system or the flow of gene expression (white bold arrow). Genetic information is expressed through transcription and translation to produce mature or refined protein. **(B)** Formation process of the first genetic system. It is considered that the genetic system was formed as going upstream against the flow of gene expression from immature or unrefined [GADV]-protein to ds-$(GNC)_n$ RNA gene. Gray bold arrows and black thin arrows indicate piling-up process of proto-members and protein synthesis process, respectively. Red bold arrows indicate maturation processes of gene and protein.

6.4. Why Could the "Mystery" of the Origin of Life Be Solved by GADV Hypothesis-Three Keys and Two Motive Forces

As described above, all of the three formation processes, the first gene, the first genetic system and the first fundamental life system, can be explained reasonably. This means that the "mystery" of the origin of life could be solved by GADV hypothesis. Furthermore, the reason, why the "mystery" could be solved by GADV hypothesis, is explained as follows.

6.4.1. Three Keys for Solving the "Mystery" of Origin of Life

There are three keys necessary to open the door for proceeding to the emergence of life. The three keys are immature [GADV]-protein, AntiC-SL RNA and ss-$(GNC)_n$ RNA, all of which were synthesized through the respective random processes and played the respective important roles in the emergence of life. For example, the first key, the immature [GADV]-proteins played roles in forming cell membrane and driving proto-metabolic systems. The second key, AntiC-SL RNA played important roles in establishing the first GNC code and forming ss-$(GNC)_n$ RNA. The third key, ss-$(GNC)_n$ RNA led to the emergence of the first life thorough formation of the first ds-$(GNC)_n$ RNA gene.

Table 1. Three keys, which led to solving the "mystery" of the origin of life.

Three keys	Synthetic method	Played key roles
1. [GADV]-peptide aggregates	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

6.4.2. Two Motive Forces Heading to the Emergence of Life

Every steps to the emergence of life should proceed under random reactions on the primitive Earth. In the situations, it would be impossible for any phenomenon to proceed to a specific direction, such as a direction of the emergence of life. However, the fact, that the first life emerged after the first ds-$(GNC)_n$ RNA gene was formed through random processes, clearly indicates that something should head to the direction of formation of the first gene and as the results the first life emerged. The motive forces were immature [GADV]-proteins and proliferating ability of [GADV]-microspheres. In

other words, synthesis of polymers such as [GADV]-polypeptides increased internal osmotic pressure in [GADV]-microspheres and induced division and proliferation of the microspheres. [GADV]-microspheres with a larger proliferating ability, which was caused by a higher catalytic activity of immature [GADV]-proteins, were selected among various [GADV]-microspheres with a different proliferation ability. Consequently, the first genetic system could be formed in the selected [GADV]-microspheres.

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

Two motive forces	Role of the motive forces
1. [GADV]-peptide aggregates	Increase of osmotic pressure by [GADV]-peptide synthesis in [GADV]-microsphere
2. [GADV]-microsphere	Selection of microspheres with a high proliferation ability headed towards the emergence of life

6.5. How Was the First Gene Expression System Formed?

It could be understood several years ago that the genetic system composed of gene, genetic code (tRNA) and protein was formed in a bottom-up manner, from protein to gene. As a matter of course, the flow of gene expression progresses from gene to protein in a top-down manner. Therefore, it is not always guaranteed that the mechanism for gene expression can be formed, if the members were piled-up onto immature proteins. Therefore, one problem, why the first gene expression system, which functions in a top-down manner, could be acquired through the processes, in which proto-members were piled-up from protein side or in a bottom-up manner, still remains unsolved.

In order to answer to the question, it is necessary to confirm the reasons, why the proto-members could be formed, as retracing in the reverse order from upstream of the gene expression flow to immature [GADV]-protein of the most downstream one by one.

Why the first ds-$(GNC)_n$ gene could be formed?

The way, how the first ds-$(GNC)_n$ gene could be formed, would be to refine a weak catalytic activity on an immature [GADV]-protein, which was produced by transcription-translation of either one strand of ds-$(GNC)_n$ RNA (Figure 7 (B)). Maturation of the immature protein became possible by using another strand of ds-$(GNC)_n$ RNA as a storage device of base substitutions.

Thus, it can be considered that the first gene encoding a mature [GADV]-protein could be generated, if ds-$(GNC)_n$ RNA could be obtained. Then, the following question, why such ds-$(GNC)_n$ RNA could be produced, arises.

Why a ds-$(GNC)_n$ RNA could be obtained just before formation of the first ds-$(GNC)_n$ gene?

The answer is because a ss-$(GNC)_n$ RNA could be obtained. Then, why could such a ss-$(GNC)_n$ RNA be obtained?

Why a ss-$(GNC)_n$ RNA could be formed before formation of the ds-$(GNC)_n$ RNA?

The reason is because a number of triplet GNCs carried by small hairpin-loop RNA (AntiC-SL RNA) composed of 17 nucleotides could be joined without gaps. Then, why could such hairpin-loop RNAs be formed?

Why hairpin-loop RNAs, which carries GNC in the loop, could be formed before formation of ss-$(GNC)_n$ RNA?

The answer is because nucleotides could be synthesized by immature [GADV]-protein catalysts and stable hairpin-loop RNAs could be formed during cycles of repeated random joining of nucleotides and degradation of oligonucleotides in [GADV]-microspheres. Then, why could nucleotides be produced in [GADV]-microspheres?

Why did nucleotides accumulate in [GADV]-microsphere before hairpin-loop RNA formation?

The reason, why nucleotide synthetic system was formed in [GADV]-microsphere, is because immature [GADV]-proteins, which drove the nucleotide synthetic system, had been confined in [GADV]-microsphere.

Why could immature [GADV]-proteins be confined in [GADV]-microsphere?

The reason is because free immature [GADV]-proteins were accidentally confined in [GADV]-microspheres, which were formed with immature [GADV]-proteins, when [GADV]-microspheres and immature [GADV]-proteins were formed by repeated wet-drying cycles of [GADV]-amino acids. Then, why immature [GADV]-proteins could be produced by random joining of [GADV]-amino acids?

Why could immature [GADV]-proteins be produced by direct random joining [GADV]-amino acids before nucleotide syntheses?

The answer is because [GADV]-amino acid composition is one of protein 0th-order structure, in which immature [GADV]-proteins could be formed by random joining of [GADV]-amino acids and because sufficient amounts of [GADV]-amino acids could be produced with prebiotic means and accumulated on the primitive Earth.

As described above, formation of [GADV]-microspheres is indispensable in order to pile-up nucleotides or ATP and CCA produced through primeval metabolic pathways onto the immature [GADV]-proteins (actually [GADV]-peptide aggregates) as a machine, which headed to the direction of formation of the first gene (Figure 6). Therefore, one of strong points of GADV hypothesis is to be able to explain reasonably the steps to the emergence of life in a bottom-up manner. Thus, there is no step to the emergence of life, which could not overcome, in GADV hypothesis. In other words, all the steps can be understood as that formation process of the genetic system could be achieved by selection of [GADV]-microspheres, in which immature [GADV]-proteins with a higher catalytic activity could be produced more efficiently than others.

7. Discussion

In this review article, it is described as focusing on the scenario for formation of the first genetic system. It is considered in the scenario, that the first gene was formed by piling-up the five members onto immature [GADV]-proteins. It is also considered that both the first genetic system and the first fundamental life system as the mechanisms, with which mature [GADV]-proteins with a high catalytic activity can be efficiently produced, could be established at the moment when the first gene was formed. The first life emerged as acquiring a number of genes necessary to live.

However, various life forms, which are different from the first life, which has lead to the modern organisms living on the present Earth, should also emerge on the primitive Earth. However, only evolutionary processes of the first life, which succeeded to evolve to the modern organisms living as using DNA genes, the universal genetic code encoding 20 amino acids and proteins composed of 20 amino acids, are presented as GADV hypothesis in this article. Therefore, for example, various nucleic acid analogues, which are different from ATP used in modern organisms as an activator of chemical reactions, must be synthesized on the primitive Earth [29-31]. Not only that, various life forms should also emerge on the primitive Earth. However, it is considered that those life forms eventually became extinct due to losing out to competition with the first life leading to the modern organisms, although such life forms using nucleic acid analogues might evolve to some extent. Therefore, all of the modern organisms are undoubtedly living on the present Earth owing to having many good lucks miraculously. The results indicate that the proto-type members, which were obtained at the respective evolutionary steps to the emergence of life, had splendid properties [21]. That is the reason why diversified organisms can live on the present Earth.

Lastly, discuss a possibility about whether or not the genetic system could be formed by starting from immature gene. Some readers of this review article who have understood the way how the first gene could be formed from immature [GADV]-proteins, may consider that the first gene could be formed even in a top-down manner, if formation of the first genetic system started from immature gene, because formation processes could be explained by exchanging mature protein for immature protein in the bottom-up manner, which GADV hypothesis supposes. However, it is necessary to understand that the genetic system never be formed from gene side or in a top-down manner, even if formation of the system began from immature genes. The reason is because immature gene never be even assumed in the absence of protein and also because formations of AntiC-SL RNA (proto-AntiC-SL tRNA), ss-$(GNC)_n$ RNA (proto-$(GNC)_n$ mRNA) and so on are indispensable to generate ds-

(GNC)$_n$ RNA sequence. Thus, only one way for formation of the first gene is to piling-up the necessary proto-members from immature [GADV]-proteins in order (Figure 7 (B)) [28].

Then, consider whether RNA, which was synthesized by random joining of nucleotides, could encode an immature protein or not. In even the case too, it is necessary to transfer genetic information on RNA strand into amino acid sequence. In the case of a singlet code, RNA having a random nucleotide sequence might encode a polypeptide chain composed of four amino acids. However, the singlet code never evolve to the universal genetic code composed of triplet code, because the singlet code became meaningless at the instance when the singlet code evolved to a doublet code and/or the doublet code evolved to a triplet code. If the genetic code started from a doublet code or from a triplet code, a machine corresponding to tRNA used in Earth lives becomes necessary to transfer the doublet code or the triplet code into amino acid sequence. However, even such machine never be formed in the absence of protein, because the correspondence between the doublet codons or the triplet codons and amino acids cannot be determined in the absence of protein.

References

1. Gilbert, W. The RNA world. *Nature*, **1986**, 319: 618.

2. Fine, J.L.; Pearlman, R.E. On the origin of life: an RNA-focused synthesis and narrative. *RNA*, **2023,** 29: 1085-1098.

3. Salditt, A.; Karr, L.; Salibi. E.; Le Vay. K.; Braun, D.; Mutschler, H. Ribozyme-mediated RNA synthesis and replication in a model Hadean microenvironment. *Nat. Commun.* **2023,** 14: 1495.

4. Oparin, A.I. *The Origin of Life on Earth*, Academic Press, New York, **1957**.

5. Fox, S.W.; Dose, K. *Molecular Evolution and the Origin of Life*, Marcel Dekker Inc., New York, USA, **1977**.

6. Kauffman, S.A. Autocatalytic Sets of Proteins. *J. Theor. Biol*. **1986**, 119: 1-24.

7. Vasas, V.; Fernando, C.; Santos, M.; Kauffman, S.; Szathmáry, E. Evolution before genes. *Biol. Direct*, **2012**, 7: 1-14.

8. Shapiro, R. The improbability of prebiotic nucleic acid synthesis. *Orig. Life*, **1984**, 14: 565-70.

9. Shapiro, R. A replicator was not involved in the origin of life. *IUBMB Life*, **2000**, 49: 173-176.

10. Shapiro R: A simpler origin for life. *Sci. Am*. **2007**, 296: 46-53.

11. Corliss, J.B.; Dymond, J.; Gordon, L.I.; Edmond, J.M.; von Herzen, R.P.; Ballard, R.D.; Green, K.; Williams, D.; Bainbridge, A.; Crane, K.; van Andel, T.H. Submarine thermal springs on the galapagos rift. *Science*. **1979**; 203: 1073–1083.

12. Holm, N.G.; 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.

13. 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.

14. Kruger, K.; Grabowski, P.J.; Zaug, A.J.; Sands, J.; Gottschling, D.E.; Cech, T. R. Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequenceof *Tetrahymena*; *Cell*, **1982**; 31: 147–157.

15. Guerrier-Takada, C.; Gardiner, K;, Marsh, T.; Pace, N.; Altman, S. The RNA moiety of ribonuclease P is catalytic subunit of the enzyme. *Cell* **1983**; 35: 849-857.

16. Crick, F. Central dogma of molecular biology. *Nature*, **1970**; 227, 561-563.

17. Joyce, G.F.; Szostak, J.W. Protocells and RNA Self-Replication. *Cold Spring Harb. Perspect Biol*. **2018**; 10: a034801.

18.   Bernhardt, H.S. The RNA world hypothesis: the worst theory of the early evolution of life (except for all the others). *Biol Direct*, **2012**; 13: 7-23.

19.   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.

20.   Ikehara. K. Pseudo-replication of [GADV]-proteins and Origin of Life, *Int. J. Mol. Sci.,* **2009**; 10: 1525-1537.

21.   Ikehara, K. *Towards Revealing the Origin of life.—Presenting the GADV Hypothesis*; Springer Nature, Gewerbestrasse: Cham, Switzerland, **2021**.

22.   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.

23.   Ikehara, K. Possible steps to the emergence of life: The [GADV]-protein world hypothesis. *Chem. Rec.,* **2005**; 5: 107–118.

24.   Dill, K.A. Dominant forces in protein folding. *Biochemistry*, **1990**; 29: 7133-7155.

25.   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.

26.   Ikehara, K. Why were [GADV]-amino acids and GNC codons selected and how was GNC primeval genetic code established? *Genes* **2023**; 14: 375.

27.   Taghavi, A.; van der Schoot, P.; Berryman, J.T. DNA partitions into triplets under tension in the presence of organic cations,

28.   with sequence evolutionary age predicting the stability of the triplet phase. *Q. Rev. Biophys.* **2017**; e15.

29.   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: https://doi.org/10.18103/mra.v12i7.5457

30.   Nielsen, P.E.; Egholm, M.; Berg, R.; Buchardt, O. Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science*, **1991**; 254: 1497-1500.

31.   Schoning, K.U.; Scholz, P.; Guntha, S.; Wu, X.; Krishnamurthy, R.; Eschenmoser, A. Chemical etiology of nucleic acid structure: the a-threofuranosyl-(3′ → 2′) oligonucleotide system. *Science*, **2000**; 290: 1347–1351.

32.   Zhang, L.; Peritz, A.; Meggers, E. A simple glycol nucleic acid. *J. Am. Chem. Soc*. **2005**, 127: 4174–4175.