THE EUROPEAN SOCIETY OF MEDICINE
Medical Research Archives, Volume 12 Issue 7

REVIEW ARTICLE

The first genetic system was established not in the top-down manner (RNA world hypothesis) but in the bottom-up manner (GADV hypothesis)

Kenji Ikehara[1,2,3]

[1]G&L Kyosei Institute, in the Keihanna Academy of Science and Culture (KASC), Keihanna interaction plaza, Lab. Wing 3F, 1-7, Hikaridai, Seika-cho, Souraku, Kyoto 619-0237, Japan.
[2]International Institute for Advanced Studies, Kizugawadai 9-3, Kizugawa, Kyoto 619-0225, Japan.
[3]Emeritus Professor, Nara Women's University, Kitauoya Higashimachi, Nara, 630-8263, Japan.

OPEN ACCESS

PUBLISHED
31 July 2024

CITATION
Ikehara, K., 2024. 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, [online] 12(7). https://doi.org/10.18103/mra.v12i7.5457

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

ISSN
2375-1924

ABSTRACT

What is necessary to elucidate the origin of life, is to make clear establishment process of the first genetic system composed of gene, tRNA (genetic code) and protein. Two ideas, RNA world hypothesis and [GADV]-protein world hypothesis or GADV hypothesis, have been proposed to elucidate the origin of life through making clear the establishment process of the genetic system. [GADV] or GADV means four amino acids; Gly [G], Ala [A], Asp [D] and Val [V]. In the RNA world hypothesis, it is considered that the genetic system was formed in order of gene, tRNA (genetic code) and protein along the flow of gene expression or in a "top-down manner". Contrary to that, the GADV hypothesis is an idea that the genetic system was formed in order of protein, tRNA (genetic code) and gene as going against the flow of gene expression or in a "bottom-up manner". It was investigated which one is valid for explaining formation process of the first genetic system. Consequently, it has been concluded that the origin of life cannot be made clear by the RNA world hypothesis or in the "top-down manner". On the other hand, it has been confirmed that the process from chemical evolution to emergence of the first life can be reasonably explained according to the GADV hypothesis or in the "bottom-up manner".

1. Introduction

THE GENETIC SYSTEM AND THE ORIGIN OF LIFE

The genetic system composed of mainly gene and protein provides the basis for modern organisms to live. Therefore, understanding, how the first genetic system was established, is one of important matters for elucidation of the origin of life. Therefore, it must be indispensable to give an answer to the question, "How was the first genetic system established?", from a standpoint of not only molecular biology but also elucidation of the origin of life. One representative of the genetic system is the "Central dogma", which was proposed by F. Crick[1] (Fig. 1. (A)). The central dogma is based on two processes in gene expression. One is transcription, in which genetic information on DNA is transferred to mRNA, and the other is translation, in which the genetic information transferred to mRNA is translated into an amino acid sequence of a mature protein with tRNA (genetic code) and ribosome (Fig. 1. (A)).

However, the role of mRNA in the Central dogma is overestimated irrespective of a similar situation, under which both DNA and mRNA play as carriers of genetic information. Consequently, the role of tRNA or genetic code in the genetic system is underestimated. Therefore, the genetic system proposed as the Central dogma should be redrawn as the Core life system[2], as shown in Fig. 1 (B). I am preferential to use the core life system as representing the genetic system.

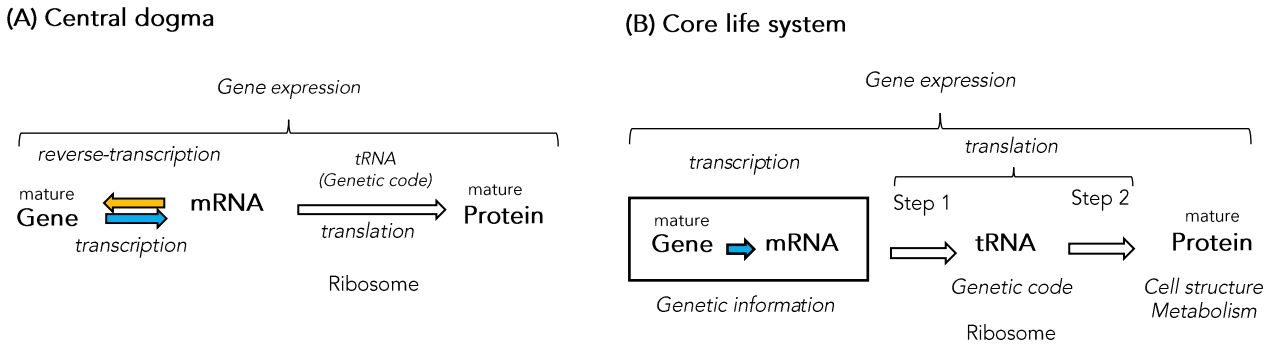

Fig. 1. Two representations, Central dogma and Core life system, of the genetic system: **(A)** Two processes for gene expression, transcription and translation, are emphasized in the Central dogma. Therefore, while mRNA is overestimated, tRNA (genetic code) is underestimated in RNA world hypothesis. (B) Gene and mRNA are collected as genetic information carriers. Consequently, tRNA and genetic code are highlighted in the Core life system.

IDEAS PRESENTED IN ORDER TO SOLVE THE MYSTERY OF THE ORIGIN OF LIFE

The following two ideas, hydrothermal vent hypothesis and space-origin theory, which have been presented for elucidation of the origin of life, are first discussed.

(1) As well known, hydrothermal vent hypothesis has been published as a possible place where the first life arose[3-5]. However, "the riddle" of the origin of life could not be solved under the hypothesis, unless the place, where a life is just emerging, is discovered, or unless traces of the lives are discovered around hydrothermal vents. The reason is because formation process of the genetic system never be made clear under the hydrothermal vent hypothesis, even if amino acids, nucleotides and other biomolecules could be produced on the hydrothermal vents.

(2) Space-origin theory is also proposed to elucidate the origin of life[6-8]. However, "the riddle" of the origin of life could not be also solved under the theory, unless newly emerged space lives are discovered on planets or satellites other than Earth. The reason is because formation process of the genetic system never be made clear based on the theory, even if amino acids, nucleotides etc. were discovered from meteorites, asteroids and so on.

Therefore, the two ideas are not discussed any more in this article, because the ideas only aim at searching the place where life emerged, but not solving the riddle of the origin of life through formation process of the genetic system.

As described above, it is necessary to make clear formation process of the genetic system in order to solve the origin of life. However, it would be generally quite difficult to make clear the formation process, because there exists the so-called "chicken and egg relationship" between gene (DNA) and protein[9,10] and, therefore, the elucidation of the origin of life falls into a circular argument.

On the other hand, one idea, RNA world hypothesis, based on which formation process of the genetic system could be explained while avoiding the "chicken and egg problem", was proposed by W. Gilbert in 1986[10] (Fig. 2). Then, consider whether or not the formation process of the genetic system can be explained by the RNA world hypothesis in the following section.

2. RNA world hypothesis

As well known, the RNA world hypothesis was proposed upon discoveries of a ribozyme, which is RNA having a catalytic activity[11,12]. Simultaneously, many persons expected that the problem of the "chicken-egg relationship" between gene and protein might be solved, because it was considered that various ribozymes having both genetic information and catalytic activity could be produced in RNA world, which was formed by amplification or self-replication of RNA[10]. It has been also expected that the genetic information and the catalytic activity on ribozymes should be transferred to DNA and protein, respectively, to complete the genetic system for the first life to emerge[10] (Fig. 1 (A)).

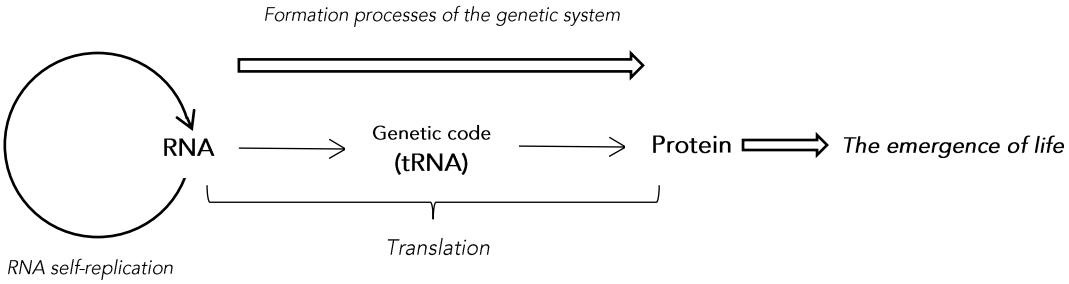

Fig. 2. Formation process of the genetic system deduced from the RNA world hypothesis. It is assumed by the RNA world hypothesis that the genetic system was formed along the flow of gene expression or in "top-down manner". That is, it is considered in the hypothesis that tRNA (genetic code) and protein were produced in order, after RNA was first formed by self-replication, for the first life to emerge.

The RNA world hypothesis is surely an interesting idea, because formation process of the genetic system might be explained by the hypothesis. However, there are many difficult obstacles to overcome in RNA world hypothesis (Fig. 3 (A)). It would be difficult to synthesize a sufficiently large amounts of nucleotides, which are necessary for life to emerge, with prebiotic means (Fig. 3 (A)-1). It would be also difficult to polymerize nucleotides to produce RNA with prebiotic means (Fig. 3 (A)-2). Furthermore, it would be quite difficult or probably impossible for RNA to self-replicate (Fig. 3 (A)-3).

Therefore, it could be understood that it is already difficult at this point to solve the origin of life according to the RNA world hypothesis. However, many investigations about the origin of life have been carried out under the RNA world hypothesis still now[13-15]. That is well understood from the fact that two review articles about the RNA world hypothesis have been recently published[16,17]. Nevertheless, there are more serious obstacles, which are almost impossible to overcome in the RNA hypothesis as follows.

2.1. CAN GENETIC INFORMATION FOR PROTEIN SYNTHESIS BE WRITTEN INTO SELF-REPLICATED RNA?

It is described in every textbook that three-dimensional structure of a protein is automatically formed, if a primary structure or an amino acid sequence of a protein encoded by a gene is given[18]. That is, a mature protein with a tertiary structure can be easily produced, once a primary structure of a protein is given by expression of a gene or one-dimensional information for protein

synthesis on DNA and/or RNA. However, it is not described anywhere how DNA and/or RNA has acquired the one-dimensional amino acid sequence information of a protein. Nevertheless, it is considered as a prerequisite that it should be possible to write genetic information or a gene into a self-replicated RNA in the RNA world hypothesis. In other words, the genetic information must be written into a self-replicated RNA in order to establish the genetic system according to the RNA world. However, it would be impossible for the self-replicated RNA even to understand in the absence of protein, the way how one-dimensional genetic information for protein synthesis could be acquired (Fig. 3 (B)-4). The origin of life has been discussed in the RNA world hypothesis without answering the question how genetic information could be written into RNA. It is only vaguely considered in the hypothesis that a self-replicated RNA should acquire genetic information for protein synthesis one day.

Next, simply consider from a standpoint of gene expression whether or not a self-replicated RNA can acquire genetic information for protein synthesis. However, such RNA having genetic information never be formed, because the RNA cannot determine a protein, of which genetic information should be written, in the absence of protein. These mean that RNA never be able to acquire genetic information of any protein in the absence of protein in principle (Fig. 3 (B)-4).

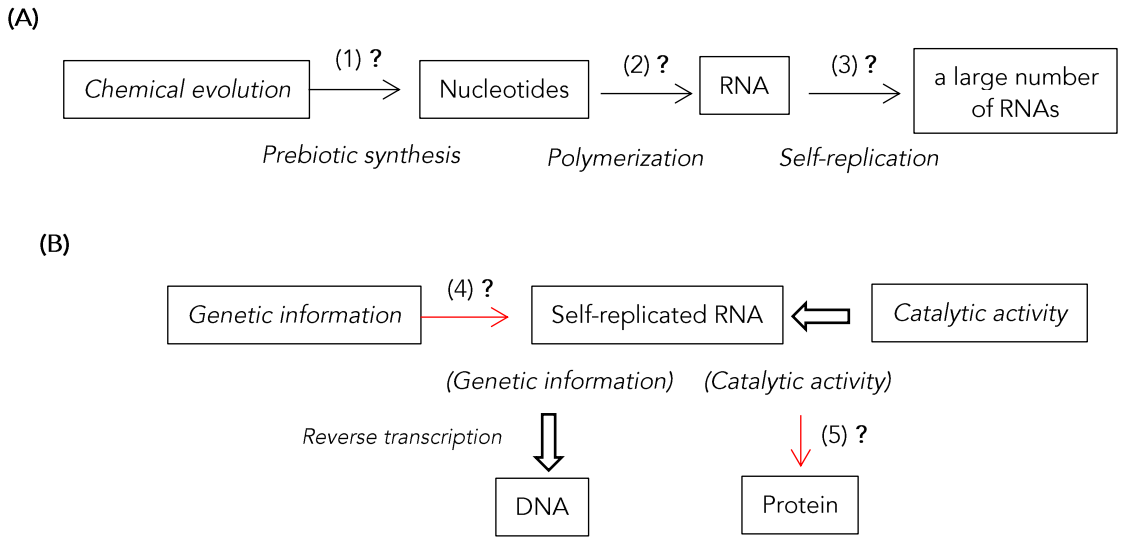

Fig. 3. (A) From chemical evolution to accumulation of self-replicated RNA. Progresses of the three evolutionary processes ((1) ~ (3)) would be difficult and might be impossible. **(B)** Acquisition of genetic information and catalytic activity by self-replicated RNA and transference of the two functions into DNA and protein, respectively. However, it would be principally impossible for self-replicated RNA to acquire the genetic information and to transfer catalytic activity onto protein, respectively (see the text; Sections 2.1 and 2.2). White bold arrows, brack arrows and red arrows indicate possible, difficult, and impossible steps, respectively.

2.2. CAN CATALYTIC ACTIVITY ON A SELF-REPLICATED RNA BE TRANSFERRED ONTO A PROTEIN?

As described above, it is considered in the RNA world hypothesis that a catalytic activity of a self-replicated RNA is transferred onto a surface of a protein one day. Considering existence of ribozymes, it could be easily supposed that self-replicated RNAs have a catalytic activity (Fig. 3 (B)). On the other hand, it becomes a problem if the catalytic activity of the self-replicated RNA can really be transferred onto a protein, because catalytic activities of ribozymes must be transferred into proteins to establish the genetic system. However, that would be impossible. The reasons are as follows.

Two strategies can be considered for transfer of a catalytic activity on a ribozyme onto a protein.

(i) The first strategy: Direct transfer of a catalytic activity of a ribozyme onto a protein

However, it would be impossible in principle to transfer catalytic activity on a ribozyme with a

three-dimensional structure directly onto a protein also having a three-dimensional structure (Fig. 3 (B)-5), because it is impossible to associate the surface structure of the catalytic RNA directly with that of a protein.

(ii) The second strategy: Indirect transfer of a catalytic activity of a ribozyme onto a protein through expression of genetic information on RNA strand.

However, it would be also impossible to transfer indirectly the catalytic activity on the RNA strand onto a protein (Fig. 3 (B)-5), because the catalytic activity of a ribozyme never be reproduced as the same catalytic activity on a surface of a protein, into which the polypeptide strand was produced and folded through the genetic information formed from RNA, and because it is impossible to make a base sequence of the catalytic RNA correspond to an amino acid sequence of a protein, even if base sequence of the RNA could be associated with an amino acid sequence of a protein.

Thus, it would be obvious that formation process of the genetic system, which is the key point for elucidation of the origin of life, cannot be explained by RNA world hypothesis based on the "gene-early theory".

3. Can the genetic system be established in top-down manner?

Then, consider whether or not it is possible to form the genetic system starting from the most upstream end of the flow of gene expression or from formation of gene to protein as expected by RNA world hypothesis, which rests on the gene-early theory.

Of course, any genetic information must be written on a self-replicated RNA for the RNA to function as a gene according to the RNA world hypothesis. However, it would be impossible to write any genetic information into self-replicated RNA as already described in detail in Section 1.1. Certainly it would be impossible to write genetic information into a self-replicated RNA. However, it becomes impossible to discuss even a possibility of formation process of the genetic system according to the RNA world hypothesis, if self-replicated RNA cannot acquire genetic information.

Then, assume in this section that self-replicated RNA could acquire the genetic information for protein synthesis in order to discuss further, whether or not formation process of the genetic information can be explained by starting from gene or in the "top-down manner".

3.1. CAN SINGLE-STRANDED (M)RNA BE GENERATED AFTER FORMATION OF DOUBLE-STRANDED RNA (GENE)?

Before considering the main point of this issue, confirm first whether or not the RNA world hypothesis really rests on the gene-early theory, because it is considered in the hypothesis that life should emerge from (m)RNA, which is placed between DNA and protein in the central dogma (Fig. 1 (A)). This may mislead sometimes when the origin of life is considered based on the central dogma, as both DNA and mRNA are regarded as a genetic carrier in the core life system (Fig. 1 (B)). It can be also understood from the fact that the base sequence of RNA can be reverse-transcribed into DNA (Fig. 1 (A)). Therefore, it is not always necessary to state that life arose from gene or genetic information, which was written on DNA.

(Top-down) RNA world hypothesis

(1) ds-RNA (gene) **// (?)** ----------**(?)** mature Protein
(2) ds-RNA (gene) ---> ss-(m)RNA ---> **// (?)** ---------- **(?)** mature Protein
(3) ds-RNA (gene) ---> ss-(m)RNA ---> (t)RNA (genetic code) ---> **//(?)** mature Protein
(4) ds-RNA gene ---> ss-mRNA ---> tRNA (genetic code) ---> mature Protein

Table 1. Formation process of the genetic system in a top-down manner deduced from RNA world hypothesis. In this case, the genetic system starts from formation of double-stranded (ds)-RNA (gene). However, it could not be understood even what should be made for formation of mature protein even if ds-RNA gene, single-stranded (ss)-mRNA and/or anticodon stem-loop (AntiC-SL) tRNA (genetic code) could be formed. //(?) indicates unknown process. Note that mature protein can be produced only after both ss-mRNA and tRNA (genetic code) were formed.

Then, consider whether or not it is possible to establish the genetic system forwards the direction from RNA to protein as considered in the "top-down manner" (Table 1). In this case, double-stranded (ds)-RNA must be first formed, although ds-RNA is only a chemical material without genetic information. It might be possible to replicate ds-RNA, because the replication could progress only with base-pair formation. However, it would be much difficult to transcribe base sequence on a ds-RNA into single-strandd (ss)-RNA, because ds-RNA could not understand the reason why ss-RNA must be synthesized through transcription especially for protein synthesis. Nevertheless, assume also here that ss-RNA could be synthesized by transcription of the ds-RNA, to move forward the discussion.

3.2. CAN TRNA BE GENERATED AFTER FORMATION OF SINGLE-STRANDED (M)RNA?

It would be impossible to generate a gene before tRNA is formed, because gene always requires tRNA to express genetic information on mRNA and, therefore, only RNA without genetic information should be produced in the absence of tRNA. Then, assume here that gene could be generated before tRNA formation, in order to further move forward the discussion, "Can tRNA be generated after formation of gene?". However, there is no way for such gene, actually only ds-RNA, to know how the genetic information can be expressed by using what kind of chemical materials. This means that it would be also impossible to generate tRNA after formation of gene (Table 1).

It would be much more difficult or probably impossible to advance translation process-1, because it is unknown even the reason why a precursor of tRNA must be formed and why protein must be synthesized as using the precursor RNA (Fig. 4). Furthermore, the reasons why the RNA must carry anticodon in the molecule and which bases are translational units, singlet, doublet or triplet codons, should be also unknown in the absence of genetic code. Therefore, it would be also impossible to advance translation step-2 (Fig. 4), because it is impossible in principle to determine the correspondence relationship between a codon or anticodon and an amino acid in the absence of protein. That is, there is also no way to know in advance how ss-RNA can be translated by using what type of chemical materials or RNA.

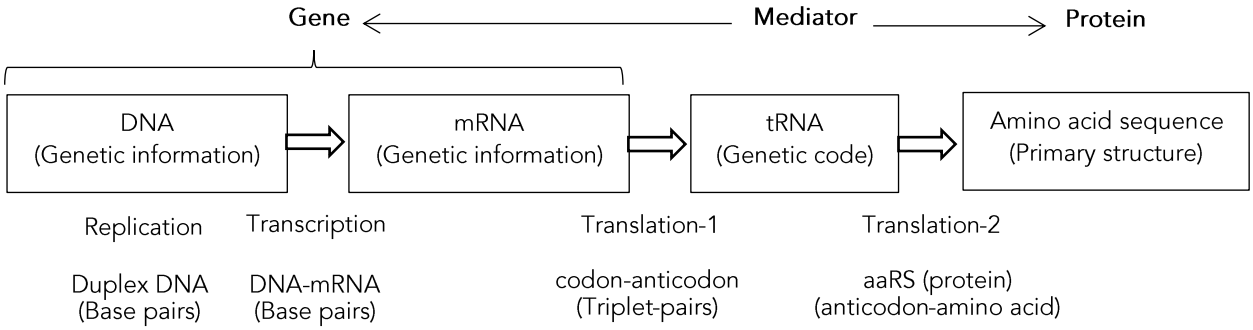

Fig. 4. Genetic system composed of DNA, mRNA, tRNA (genetic code) and protein: It is discussed in this review article, whether or not how the system was established towards from upstream to downstream or in the "top-down manner" or from downstream to upstream or in the "bottom-up manner".

3.3. CAN PROTEIN BE PRODUCED AFTER FORMATION OF TRNA?

Next, consider whether or not tRNA, which is necessary to synthesize protein, can be formed before formation of protein. Only meaningless RNA, but not tRNA, might be formed independently of protein. However, it would be also impossible for a precursor RNA of tRNA to know the way how to transmit genetic information on a gene for synthesis of a polypeptide chain to produce a functional protein (Fig. 4). The reason is because there is no way to know how protein can be synthesized with the RNA.

Therefore, gene, actually ds-RNA, which was formed independently of protein and tRNA, must be meaningless, because it is impossible to know the way how protein can be produced by using what type of RNA (Fig. 4).

3.4. CAN A POLYPEPTIDE CHAIN SYNTHESIZED IN A "TOP-DOWN MANNER" BE FOLDED INTO A TERTIARY STRUCTURE WITH A CATALYTIC ACTIVITY?

Nevertheless, assume again here that a polypeptide chain could be synthesized with a self-replicated RNA. It is well known that secondary and tertiary structures of a protein are automatically formed, if a primary structure is given by a corresponding gene (Fig. 5). However, the polypeptide chain produced by a self-replicated RNA, which has not genetic information, never be folded into a rigid tertiary structure like as a mature protein, because the polypeptide chain cannot understand how the chain should be folded into what a three-dimensional structure and what a catalytic function is exhibited by the structure folded into a three-dimensional structure. This means that it is impossible to produce a mature protein with a necessary catalytic activity for a life to live in the "top-down manner", because there is no way to know how the polypeptide chain is folded into what type of three dimensional structure, which exhibits what type of catalytic activity (Fig. 5).

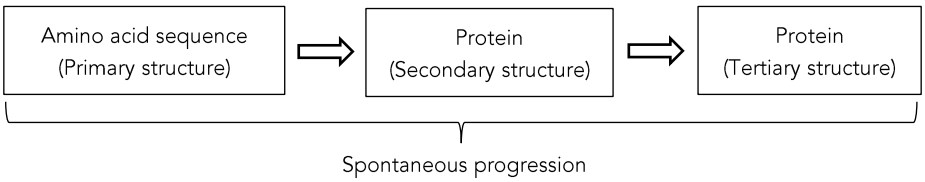

Fig. 5. Tertiary structure formation of protein: It is well known that a polypepyide chain or primary structure synthesized under gene expression is spontaneously folded as assembling secondary structures into a protein with a tertiary structure. This means that a mature protein must be formed as adjusting the tertiary structure of the immature protein through amino acid replacements. Therefore, this also indicates that a mature protein must be always formed from an immature protein in the "bottom-up manner".

3.5. WHY HAS THE FORMATION PROCESS OF GENETIC INFORMATION BEEN CONSIDERED SO FAR FROM A STANDPOINT OF TOP-DOWN IN THE RNA WORLD HYPOTHESIS?

Probably, researchers studying on the origin of life, who do know only mature proteins, would consider that gene must be first formed along the flow of gene expression, because any mature protein never be produced without gene. Certainly, it would be impossible to synthesize a mature protein in the absence of gene (Fig. 4), because the probability for synthesis of even a small protein composed of only 100 amino acids by random joining of the amino acids reaches $1/20^{100} = \sim 10^{-130}$ [19]. Most researchers studying on the origin of life would consider as a prerequisite that any mature protein must be produced under gene expression.

In addition, many researchers, who give too much weight to self-replication, may expect that RNA should acquire genetic information for protein synthesis one day, if diverse self-replicated RNAs could be formed and some RNAs could be selected among the diverse RNAs. In fact, the idea has been presented in two recent review articles that the riddle on the origin of life should be solved according to the gene-early theory or the RNA world theory[16,17].

4. [GADV]-protein world hypothesis (GADV hypothesis)

It becomes necessary to reveal newly the origin of life with a different idea from the RNA world hypothesis, which is based on formation process of the genetic system. The idea is [GADV]-protein world hypothesis or GADV hypothesis, which I have proposed[20,21]. It is examined in this section whether or not it is possible to explain the steps from chemical evolution to the emergence of life or formation process of the genetic system according to the GADV hypothesis.

4.1. HOW WAS GADV HYPOTHESIS PROPOSED?

Here, it is first explained how the GADV hypothesis was proposed, because it is important to understand the reason, why I could hit upon the GADV hypothesis, which is a totally opposite idea to the RNA world hypothesis, and because it should become easily to understand the steps to the emergence of life, which the GADV hypothesis has been reaches.

The study on the GADV hypothesis did not start aiming at elucidation of the origin of life but was triggered by a study on the origin of current genes, or "How have entirely new genes been formed in modern microorganisms?", if such entirely new genes have been generated still now. Consequently, first, (1) GC-NSF(a) hypothesis, which explains that current genes have been formed from nonstop frame on antisense sequence of GC-rich gene, was acquired[22]. Successively, (2) SNS primitive genetic code hypothesis was found, suggesting that one of old genetic codes was SNS code (S and N mean G or C and either of four bases, U, C, A and G, respectively)[23], and (3) GNC primeval genetic code hypothesis, advocating that the first genetic code was GNC code, which was composed of four [GADV]-amino acids and four GNC codons, could be presented[24,25]. Finally, (4) the GADV hypothesis, assuming that the first life arose from [GADV]-protein world[20,21], which was formed by pseudo-replication of [GADV]-proteins[20,21,26], was suddenly hit upon and was presented about 20 years ago (Fig. 6).

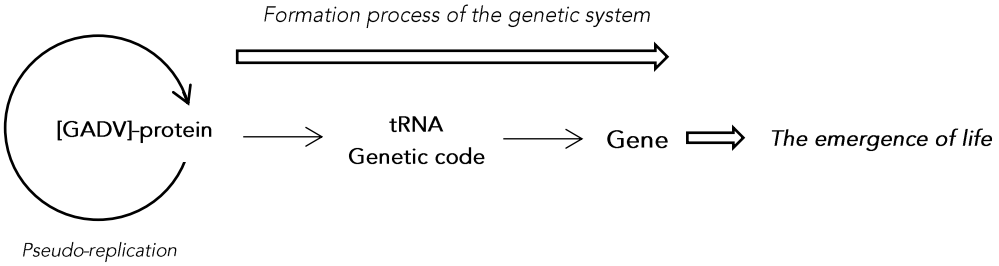

Fig. 6. The [GADV]-protein world hypothesis (GADV hypothesis): It is considered in the hypothesis that formation of the first genetic system, started from production of immature [GADV]-proteins by random joining of [GADV]-amino acids, and, was followed by formations of primeval anticodon stem-loop tRNAs (AntiC-SL tRNAs), GNC primeval genetic code and primeval double-stranded $(GNC)_n$ RNA genes as pilling up on the immature [GADV]-proteins in order. Thus, it is assumed that the genetic system was formed as going upstream against the flow of gene expression or in "bottom-up manner".

Thus, two ideas, the RNA world hypothesis and the GADV hypothesis, are directly opposite about formation process of the genetic system as the former is trying to solve the origin of life in "top-down manner" and the latter is aiming at elucidating the origin of life in "bottom-up manner" (Figs. 2 and 6). Therefore, the purpose of this review article is to confirm which idea can explain the formation process of the genetic system and can solve the "mystery", on how the first life emerged on the primitive Earth. In other words, which idea is correct, "the first genetic system was formed along the flow of gene expression as considered in the RNA world hypothesis (in a "top-down manner")" or "the genetic system was generated as going upstream against the flow of gene expression as presented by the GADV hypothesis (in a "bottom-up manner")".

5. Can the formation process of the genetic system be explained according to GADV hypothesis or in a "bottom-up manner"?

Then, consider whether the formation process of the genetic system can be explained by GADV hypothesis or not. It is considered that the acquisition process of genetic information for mature protein synthesis could be made clear owing to one of protein 0^{th}-order structures or [GADV]-amino acid composition and maturation of an immature [GADV]-protein, which was produced in the protein 0^{th}-order structure[2,20,21] (Fig. 7). Therefore, the mystery of the origin of life could be solved according to the GADV hypothesis. The reasons are explained in detail below.

5.1. FEATURES OF GADV HYPOTHESIS

GADV hypothesis is an idea based on a protein 0^{th}-order structure or the special [GADV]-amino acid composition[27]. In other words, immature but water-soluble globular [GADV]-proteins with a large flexibility could be produced by random joining of [GADV]-amino acids. Notification of the pseudo-replication of immature [GADV]-proteins or immature [GADV]-protein synthesis by immature [GADV]-proteins triggered to finding of the protein 0^{th}-order structure. Pseudo-replication and protein 0^{th}-order structure give rise to the GADV hypothesis[21,22].

5.2. STEPS TO THE EMERGENCE OF LIFE DEDUCED FROM GADV HYPOTHESIS

In the GADV hypothesis, steps from chemical evolution to the emergence of life can be explained through the respective origins of six members (protein, cell structure, metabolism, tRNA, genetic code and gene), of which the fundamental life system is composed[2] (Fig. 7).

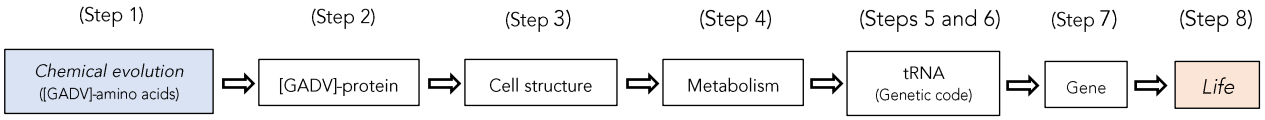

Fig. 7. Steps from chemical evolution to the emergence of life, which is deduced from GADV hypothesis[2]: The hypothesis suggests that life arose as going upstream against the flow of gene expression.

6. Can the genetic system be established in a "bottom-up manner"?

According to the GADV hypothesis, it is reasonably explained that the genetic system was established as going upstream from protein to gene against the flow of gene expression or in the "bottom-up manner" (Fig. 7). It could be well understood that a gene encoding a mature protein and other members necessary to transcribe and translate genetic information must be prepared in advance in order to produce a mature protein with a compact and rigid tertiary structure (Fig.1 (B)). Therefore, it is confirmed in this section whether or not the formation process surely progressed in the "bottom-up manner".

The word, "protein" would naturally mean a "complete and mature protein", which is encoded by a gene. As a matter of course, such a mature protein is produced only under genetic system. However, I noticed pseudo-replication, by which immature water-soluble globular proteins can be produced even by random joining of [GADV]-amino acids in the protein 0th-order structure or [GADV]-amino acid composition[20,21,25-27] (Fig. 8). The idea, that proteins could be generated even in the absence of gene, if the proteins were such immature proteins or baby proteins, came to my mind about 20 years ago[21,22]. Thus, the formation process of the genetic system in GADV hypothesis started from the immature protein production in the absence of all five members other than protein (Fig. 7).

Thus, it can be explained in GADV hypothesis that the first genetic system was established when the evolutionary process finally reached to the formation of gene, after going upstream against the flow of gene expression from the downstream-most end of immature [GADV]-protein synthesis as adding in order of cell structure, metabolism, tRNA, genetic code and gene one by one (Fig. 7).

This is the most reasonable and possible formation process of the first genetic system, as further explained below.

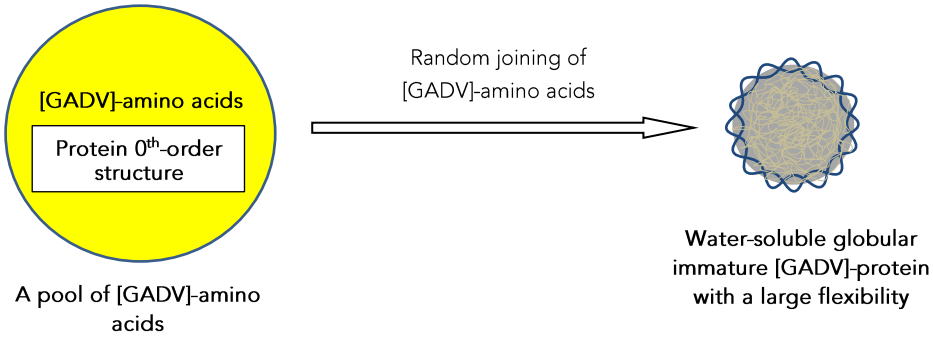

Fig. 8. Protein 0th-order structure: Water-soluble globular immature [GADV]-protein with a large flexibility or a baby protein can be produced by random joining of [GADV]-amino acids in protein 0th-order structure or a special amino acid composition, which satisfies the four conditions for water-soluble globular protein formation.

6.1. FORMATION OF THE GENETIC SYSTEM STARTED FROM IMMATURE [GADV]-PROTEINS OR IN THE "BOTTOM-UP MANNER" (*PHASE 1: STEPS TOWARDS THE FIRST GENE*)

First, it is explained that the reason, why the genetic system must be formed from protein but not from gene or tRNA. Many persons including researchers studying on the origin of life may consider that protein cannot be produced in the absence of gene. That is of course correct, if the protein is a mature protein. On the contrary, immature proteins, which can be synthesized under

a protein 0^{th}-order structure, could be produced in the absence of gene and tRNA, as considered in GADV hypothesis (Fig. 8). Another reason is because proteins are working polymers and, therefore, proteins themselves are meaningful in the absence of gene and tRNA. Therefore, it is concluded that only immature proteins can be produced owing to the protein 0^{th}-order structure even in the absence of gene and tRNA (Fig. 7 and Table 2). That is, it is possible to form protein independently of gene and tRNA, if the protein is an immature protein.

6.2. FORMATION OF THE GENETIC SYSTEM WAS SUCCEEDED BY AN RNA AFTER IMMATURE [GADV]-PROTEIN PRODUCTION

Furthermore, it would be possible to produce an anticodon stem-loop (AntiC-SL) RNA[2], but not tRNA at this time point, or even in the absence of gene in order to facilitate immature protein synthesis (Table 2). Namely, the AntiC-SL RNAs as a part of activators were formed to realize more efficient synthesis of immature [GADV]-proteins than direct joining of [GADV]-amino acids. Such activators are shown in Table 2 (Phase 1) and below.

Phase 1 to the first gene:

(1) Use of [GADV]-amino acids activated with ATP, or [GADV]-aa-AMP.

(2) Use of [GADV]-amino acids activated with $^{5'}$-CCA-$^{3'}$, or [GADV]-aa-$^{3'}$-ACC-$^{5'}$.

(3) Use of [GADV]-amino acids activated with $^{3'}$-ACC-$^{5'}$-AntiC-SL RNA, or [GADV]-aa-$^{3'}$-ACC-$^{5'}$-AntiC-SL RNA.

(4) Dimer formation of [GADV]-aa $^{3'}$-ACC-$^{5'}$AntiC-SL RNA without GNC codon, which were juxtaposed side-by-side, accelerated immature [GADV]-protein synthesis, owing to proximity effect of two [GADV]-amino acids, which were bound with ACC-AntiC-SL RNAs.

(5) Tetramer of [GADV]-aa-$^{3'}$-ACC-$^{5'}$-AntiC-SL RNA with GNC codon, which were formed by vertical binding of two juxtaposed dimers, accelerated further the immature [GADV]-protein synthesis. Thus, proto-AntiC-SL tRNA was formed as one of activators for immature [GADV]-protein synthesis.

Phase 1 to the first gene:

immature [GADV]-protein synthesis by direct joining of [GADV]-amino acids
immature [GADV]-protein synthesis with activated [GADV]-amino acids
((1) ATP, (2) ACC, (3) ACC-AntiC-SL RNA monomer, (4) ACC-AntiC-SL RNA dimer, (5) ACC-AntiC-SL RNA tetramer)

Phase 2 to the first gene:

AntiC-SL RNA tetramer (tRNA) → immature [GADV]-protein
ss-$(GNC)_n$ RNA (mRNA) → AntiC-SL tRNA (GNC code) → immature [GADV]-protein
ds-$(GNC)_n$ RNA → ss-$(GNC)_n$ mRNA → AntiC-SL tRNA (GNC code) → mature [GADV]-protein

---

Table 2. Formation process of the genetic system in a bottom-up manner deduced from GADV hypothesis. If immature [GADV]-proteins could be synthesized more efficiently by participation of AntiC-SL tRNA and/or GNC code than synthesis of immature [GADV]-protein synthesis by direct random joining of [GADV]-amino acids, AntiC-SL tRNA and/or GNC code were added into the synthetic system of immature [GADV]-proteins. Similarly, if participation of ss-$(GNC)_n$ RNA and ds-$(GNC)_n$ RNA to the synthetic system of immature [GADV]-proteins could be further accelerated, those RNA were added to the synthetic system of immature proteins.

6.3. FORMATION OF THE FIRST GENE ENCODING A MATURE [GADV]-PROTEIN, WHICH IS THE LAST MEMBER COMPOSING THE GENETIC SYSTEM (*PHASE 2: STEPS TOWARDS THE FIRST GENE*)

The following steps must be taken to form the first gene[28] (Table 2 (Phase 2)). ss-RNA must be first formed by concatenation of anticodons carried by [GADV]-aa-$^{3'}$-ACC-$^{5'}$-AntiC-SL RNA tetramers to produce more efficiently immature [GADV]-proteins than before. Note that the ss-RNA, which was composed of a random $(GNC)_n$ codon sequence, was a proto-mRNA. Further, ds-RNA was formed by complementary strand synthesis of the ss-RNA. As a matter of course, the ds-RNA did not have genetic information for synthesis of a mature [GADV]-protein. Therefore, the problem, which remains unsolved, is the way how genetic information for synthesis of a [GADV]-mature protein can be written as a $(GNC)_n$ codon sequence into ds-RNA. Only one way for solving the problem is as follows (Fig. 9).

(1) The process started from the ds-RNA composed of random $(GNC)_n$ codon sequences encoding a random [GADV]-amino acid sequence or an equivalent to the [GADV]-amino acid sequence, which was formed by direct random joining of [GADV]-amino acids in protein 0^{th}-order structure, under which an immature water-soluble globular protein can be produced.

(2) Next, an immature [GADV]-protein having a random [GADV]-amino acid sequence was synthesized by expression of either one of $(GNC)_n$ codon sequences of the ds-$(GNC)_n$ RNA.

(3) A weak catalytic activity on the immature [GADV]-protein was gradually maturated to a more complete [GADV]-protein with a higher catalytic activity through accumulation of appropriate base substitutions on the ds-$(GNC)_n$ RNA.

(4) Thus, the first ds-$(GNC)_n$ RNA gene was finally obtained, when the immature [GADV]-protein acquired a sufficiently high catalytic activity (Fig. 9).

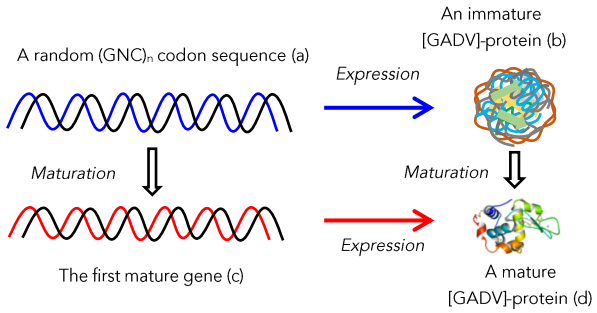

Fig. 9. Formation process of the first gene encoding the first mature [GADV]-protein: The first mature gene (c), on which a genetic information for synthesis of a mature [GADV]-protein (d), could be obtained by accumulating appropriate mutations on the random $(GNC)_n$ codon sequence or a proto-gene (a) and by maturation from a weak catalytic activity appeared on the immature [GADV]-protein (b) to a higher catalytic activity.

These are the steps proceeding from random $(GNC)_n$ RNA encoding an immature protein to acquisition of genetic information, which is exactly followed in the "bottom-up manner" (Table 2 (Phase 2)). There would be no way besides the process along which one gene encoding a mature protein can be obtained.

6.4. WHY MUST THE GENETIC SYSTEM BE FORMED IN A "BOTTOM-UP MANNER"?

The reason, why the genetic system must be formed in the "bottom-up manner", is because [GADV]-microspheres, which acquired higher proliferation abilities based on more efficient synthesis of immature [GADV]-proteins with a higher catalytic activity than before, were selected. The higher proliferation abilities were acquired through formation of six members in order of protein, cell structure, metabolism, tRNA, genetic code and gene as going upstream against the flow of gene expression (Figs. 6 and 7). In other words, acquisition of such a higher catalytic activity and a more efficient synthesis of immature [GADV]-proteins made it possible to progress to the direction of the emergence of life during the random reactions on the primitive Earth. It must be noted that there were two motive forces, immature [GADV]-proteins and [GADV]-microspheres, which moved towards the emergence of life as expected in the GADV hypothesis. On the contrary, such motive forces could not be acquired in the RNA world hypothesis. This also must be critical defect in the RNA world hypothesis.

When an airplane was manufactured for the first time, performance of a prototype 1 should be examined. Performance of a prototype 2, which was rebuilt as the blue print for the first machine was changed, based on the result of the first prototype, would be also examined. The first consummate airplane, which can fly well, must be manufactured as such works were repeated. Similarly, formation of a mature protein, the object of the genetic system, must start from an immature protein in order to establish the genetic system. Thus, it is impossible to generate genetic information for protein synthesis in the absence of protein.

7. Discussion

Two ideas, RNA world hypothesis and GADV hypothesis, have been proposed to solve the mystery of the origin of life from a standpoint of formation process of the genetic system. The former is considering the formation process from gene to protein or in the "top-down manner". On the contrary, the latter is pursuing the mystery in the "bottom-up manner". It is discussed in this review article, which idea is valid for solving the mystery. Consequently, it was found that there exist quite difficult or probably impossible steps to overcome in the RNA hypothesis, considering the formation process from gene to protein in the "top-down manner". On the contrary, it was confirmed that there is no obstacle, which cannot be overcome, in the GADV hypothesis or in "bottom-up manner". In addition, it could be confirmed that the formation process could progress to the direction of the emergence of life or from protein to gene, because of two motive forces, enhancements of catalytic activity of immature [GADV]-protein and proliferation ability of [GADV]-microsphere, which constrain to move the process towards the most upstream or the formation of the first gene. Furthermore, the reason, why the genetic system must be formed in the "bottom-up manner", could be also understood by the two motive forces.

As can be seen in Fig. 4, gene expression (phase 1) proceeds as interlocking closely at the respective steps. For examples, genetic information is transmitted through base pairs at DNA replication and transcription. Base pairs between codon-anticodon are also used to transmit genetic information from mRNA to tRNA (Fig. 4. translation -1). At translation-2 in Fig. 4, the genetic information, which was received from mRNA, is converted to amino acid sequence by tRNAs to synthesize an amino acid sequence of a mature protein.

Thus, genetic information is transferred from upstream to downstream, as indicated by each correspondence relationship like a "key and key hole" at the respective steps. On the other hand, the most meaningful matter as lives is not gene at the most upstream but working polymer, protein, at the most downstream. That could be well understood by the facts that effect of a base replacement on a gene always appears as functional change of a protein or phenotype. This means that genetic information exists only for production of a protein and, needless to state, a gene itself does not work at an actual site. Therefore, it could be also well understood that the respective steps of the gene expression, which are determined by the correspondence relationship, must be formed by piling up from the most downstream or working polymers, proteins, one by one as constructing the respective correspondence relationships and as confirming if immature [GADV]-proteins could acquire a higher catalytic activity and/or could be synthesized more efficiently than before.

On the contrary, any motive force, which can advance towards establishment of the genetic system, could not be acquired in the absence of protein or working polymer. This is the reason why the mystery on the origin of life could not be solved under the RNA world hypothesis or in the top-down manner.

8. Conclusion

The study on the origin of life has been carried out mainly according to the "gene/replicator-early theory" or RNA world hypothesis, under which it is

considered that the first genetic system was established along with the stream of gene expression, that is, from gene to protein or in a top-down manner. However, the formation process of the first genetic system has not been explained under the RNA world hypothesis still now, On the contrary, it has been confirmed that the formation process of the first genetic system can be reasonably explained according to the protein/metabolism-early theory or GADV hypothesis, under which it is considered that the first genetic system was formed as going upstream against the flow of gene expression, that is, from protein to gene or in bottom-up manner.

Conflict of Interest:

The authors declare no conflict of interest.

Funding:

This research received no external funding.

Acknowledgements:

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

References:

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

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

3. 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.

4. 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.

5. 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.

6. Arrhenius S. Evolution of the universe. 1908; Harper, London, United Kingdom.

7. Temple R. The prehistory of panspermia: astrophysical or metaphysical? 2007; Cambridge University Press, Cambridge, United Kingdom.

8. 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.

9. Sankaran N. How the discovery of ribozymes cast RNA in the roles of both chicken and egg in origin-of-life theories. *Stud Hist Philos Biol Biomed Sci.* 2012;43:741-750.

10. Gilbert W. The RNA world. *Nature.* 1986;319:618.

11. Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, Cech TR. Self-splicing RNA: autoexcision and autocyclization of ribosomal RNA intervening sequence of *Tetrahymena*. *Cell.* 1982;31:147-157.

12. 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.

13. Tagami S, Li P. The origin of life: RNA and protein co-evolution on the ancient Earth. *Dev Growth Differ.* 2023;65:167-174.

14. Müller F, Escobar L, Xu F, Węgrzyn E, Nainytė M, Amatov T, Chan CY, Pichler A, Carell T. A prebiotically plausible scenario of an RNA-peptide world. *Nature.* 2022;605:279-284.

15. Rivera-Madrinan F, Di Iorio K, Higgs PG. Rolling Circles as a Means of Encoding Genes in the RNA World. *Life.* 2022;12(9): 1373.

16. Saito H. The RNA world 'hypothesis'. *Nat Rev Mol Cell Biol.* 2022;23:582.

17. Fine JL, Pearlman RE. On the origin of life: an RNA-focused synthesis and narrative. *RNA.* 2023;29:1085–1098.

18. Berg JM, Tymoczko JL, Stryer L. *Biochemistry* 5th ed. 2002; W. H. Freeman and Company/New York.

19. Dill KA. Dominant forces in protein folding. *Biochemistry.* 1990;29:7133-7155.

20. 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.

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

22. 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.

23. Ikehara K, Yoshida S. SNS hypothesis on the origin of the genetic code. *Viva Origino.* 1996;26:301–310.

24. 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.

25. Ikehara K. Why Were [GADV]-amino Acids and GNC Codons Selected and How Was GNC Primeval Genetic Code Established? *Genes.* 2023;*14*(2):375

26. Ikehara K. Pseudo-replication of [GADV]-proteins and origin of life. *Int J Mol Sci*. 2009;10:1525–1537.

27. Ikehara K. Protein ordered sequences are formed by random joining of amino acids in protein 0^{th}-order structure followed by evolutionary process. *Orig Life Evol Biosph*. 2014;44:279–281.

28. Ikehara K. How was the first gene formed in the absence of gene. *Preprints*. 2023; 2023071624.