

LETTERS TO THE EDITORS

PACS numbers: **87.14. – g**, **87.18. – h**, 87.23.kg

Life originated as a genetic code

(contribution to the discussion of the paper by G R Ivanitskii "21st century: what is life from the perspective of physics" [*Usp. Fiz. Nauk*. **180** 337 (2010); *Phys. Usp*. **53** 327 (2010); *Usp. Fiz. Nauk*. **182** 1235 (2012); *Phys. Usp*. **55** 1152 (2012); *Usp. Fiz. Nauk* **182** 1238 (2012); *Phys. Usp*. **55** 1155 (2012)])

Yu N Korystov

DOI: https://doi.org/10.3367/UFNe.2016.11.037996

Contents

Abstract. This letter is aimed at concretizing the origin-of-life scenario. The author defines life as a self-reproducing information system and briefly discusses the current status of the origin-of-life problem. The first bimolecular origin-of-life scenario is proposed, which is based on the interaction between amino acids and triplets of nucleotides corresponding to their codons. In this scenario, the reproduction of information occurred by hydrogen bonding with complementary nucleotides and, hence, was due to the formation of complementary codons that reproduced the primary codon. The bimolecular scenario fundamentally necessitates the existence of complementary codons for all those of amino acids associated with life. A complete correspondence is shown to exist between codons and complementary codons for 21 amino acids, a fact which greatly favors the proposed hypothesis, because this correspondence is impossible if amino acids choose codons at random.

Keywords: amino acids, origin of life, genetic code

Yu N Korystov Institute of Theoretical and Experimental Biophysics, Russian Academy of Sciences,
ul. Institutskaya 3, 142290 Pushchino, Moscow region, Russian Federation
E-mail: ykorystov@rambler.ru

Received 23 August 2016, revised 17 November 2016
Uspekhi Fizicheskikh Nauk **187** (2) 235 – 240 (2017)
DOI: https://doi.org/10.3367/UFNr.2016.11.037996
Translated by Yu V Morozov; edited by A Radzig

1. Introduction

The paper by G R Ivanitskii [1] describes the origins of life as the interplay between various parts of matter, in which one part of it "acquires the ability to remember the success (or failure) probabilities from the previous rounds of the game, thereby increasing its chances for further survival in the next round." Commenting on this article, V I Klyatskin [2] expresses an opinion at variance with that held by the author of Ref. [1] and concludes that the origin of life is rather an event that occurred with the probability of unity under conditions on early Earth. In reply to the critique, Ivanitskii argues that "the main riddle of the origin of life is not the fact of structure formation itself but the time intervals from cluster formation till disappearance, i.e., cluster lifetime τ. A newly formed cluster must remember conditions of its birth to become a 'living structure' and transfer this memory to the progeny" [3]. The present paper proposes a concrete origin-of-life scenario in which the presence of the four canonical nucleotides and amino acids creates, with the probability of unity, conditions for the formation of all possible nucleotide triplets and retention (reproduction) of only those with which amino acids interact.

2. Definition of life

The definition of life is an indispensable prerequisite for the discussion of this phenomenon, for it is hardly possible to consider the origin of an event without understanding its nature and essence. Therefore, the notion of life needs to be properly defined prior to consideration of a hypothesis on its

origin. There are currently a large number of relevant definitions of life based on physiological, metabolic, biochemical, genetic, and thermodynamic grounds. However, none of them encompasses life in all its manifestations, and besides some allow the inclusion of indubitably nonliving systems [4]. In my opinion, M Eigen proposed a definition of life closest to the truthful understanding of this phenomenon [5, 6]. It reads: ''All reactions in a living system follow a controlled programme operated from an information centre, whose aim is the self-reproduction of the programme itself.'' Only one amendment to this definition is needed; specifically, the program is an algorithm (instruction) of structures that controls reactions indirectly, via the structures. The following key aspects of the essence of life can be distinguished according to Eigen's concept: (1) reproduction of information; (2) the existence of structures and reactions for this purpose, and (3) information as an algorithm (instruction) for the structures and reactions reproducing it. Let us denote the totality of structures and reactions reproducing information by the term 'operator'. Bearing in mind the aforementioned essential characteristics of life, its definition can be formulated as follows: life is a self-reproducing information system (SIS) with information (an operator structure algorithm) and the information-reproducing operator as indispensable elements. To distinguish life from other known SISs, e.g., a culture for which human brain neurons serve as material information carriers, a comprehensive definition of life must contain a reference to a specific material information carrier and an operator. Thus, the complete definition of life can be introduced in the following wording: *life is a self-reproducing information system* (*SIS*) *having information* (*an operator structure algorithm*) *materialized by a sequence of nitrogenous bases and an information-reproducing operator composed of amino acids as its indispensable elements*.

3. Current status of the origin-of-life problem

Today, most authors agree that the reproduction of information was of paramount importance for the emergence of life. Two groups into which the hypotheses about the origin of life can be subdivided differ only in what came first: metabolism ('metabolism first') or information ('genetics first') [7–9]. The discovery of RNA catalytic activity provided a basis for the hypothesis that life had begun in the form of self-reproducing RNA molecules capable of catalyzing chemical reactions, 'the RNA world hypothesis' [10–12]. However, the primordial soup is believed to have contained a large amount of nucleotides, besides canonical ones, which made any spontaneous emergence of a self-replicating RNA (composed of at least 40 solely canonical nucleotides) very unlikely [3, 13]. Moreover, such an origin-of-life scenario is fraught with unresolvable problems related to the formation of the genetic code, translation, and the involvement of amino acids and proteins in genetic programs. Information and communication technology experts have only recently come to an understanding of the logical gulf separating origin-of-life hypotheses and existing life [14]. Here are a few excerpts from Ref. [14]: ''the lack of separation between algorithm and its implementation implies that monomolecular systems are divided from known life by a logical and organizational chasm that cannot be crossed by mere complexification of passive hardware.'' And further: ''For this rather deep reason, it may be that life had to be 'bimolecular' from the start.'' The authors also emphasize that all the above hypotheses of the origins of life disregard an inherent property of biological information, i.e., its being an algorithm or instruction. (It should be added on a personal note that it is the operator instruction.) It is pointed out for the first time in the literature on the origin of life that information should be not only self-reproducing (as in the RNA world) but also meaningful (semantic). The authors say nothing about the primary information algorithm. The present publication deals with a bimolecular variant of life origin according to which amino acids formed the primary information content; also discussed is the information reproduction mechanism.

4. Bimolecular scenario of the origin of life

It is supposed that L-amino acids interacted with spontaneously formed triplets of canonical nucleotides corresponding to their codons. The amino acid–codon interaction increased the latter's lifetime by a factor sufficient for the formation of hydrogen bonds with complementary canonical nucleotides. Base stacking in complementary nucleotides bound by hydrogen bands resulted in their polymerization [15] into a complementary codon. Information contained in an initial codon was reproduced when another amino acid from the same pool stabilized the complementary codon that reproduced the initial one by forming hydrogen bonds with complementary nucleotides. In this way, the system reproduced (indirectly rather than directly) information materialized as a sequence of nitrogenous bases with the help of operators (amino acids). Information content was the amino acid operator structure. In terms of the above definition of life, such a system was a living one.

The hypothesis proposed in this article is based on the following three premises:

(1) Prebiotic conditions gave rise to nitrogenous bases and triplets of canonical nucleotides.

(2) The nitrogenous bases were stacked on top of one another and formed hydrogen bonds with complementary bases.

(3) L-amino acids interacted with their codons so that the codon lifetime increased by a factor sufficient for the formation of hydrogen bonds with three complementary canonical nucleotides, i.e., for reproduction of information via a complementary codon.

It was shown long ago that UV irradiation of dilute adenine, ribose, and phosphate solutions produces ATP [16]. Reference [17] dated to 2009 gave evidence that cyanoacetylene, cyanamide, glyceraldehyde, glycolaldehyde, and inorganic phosphate are plausible prebiotic feedstock molecules for the synthesis of pyrimidine nucleotides. Furthermore, UV radiation transformed part of dissolved cytosine into uracil. The authors of Refs [18, 19] demonstrated that the addition of amino acids containing a small excess (1%) of L-stereoisomers to the mixture used in Ref. [17] produced chirally pure ribonucleotides containing only D-ribose. Moreover, nitrogenous bases and amino acids can be obtained by heating formamide, while cyclic nucleotides are formed in the presence of a phosphate source and can undergo polymerization into oligonucleotides of various lengths through stacking interactions of their bases [15].

The validity of the second premise is universally recognized. One recent publication [20] demonstrates that stacking between canonical bases and the formation of hydrogen bonds with complementary bases occur at medium neutral

pH, because the bases bear no charge under such conditions and interact via stacking and the formation of hydrogen bonds. The author of the cited study argues that it is for this reason that canonical nitrogenous bases were chosen as life information carriers. Such interactions are equally characteristic of nitrogenous bases contained in nucleotides [21].

The hypothesis viewing the genetic code as essentially based on the interaction between amino acids and nucleotide sequences dates back to the sixties [22, 23]. Let us discuss in broad terms what this specific interaction must have been like and how it can be verified in experiment. If the genetic code is assumed to have formed via the interaction of amino acids with stacked bases inside a codon, its triplet character for all amino acids suggests that this property is due to the functional groups common to all amino acids. The only such groups are amino and carboxyl groups. Consequently, the triplet nature of the genetic code gives evidence that amino acids interacted with base triplets via these groups. Specificity of the genetic code must be determined by the R-groups of amino acids, because they differ only in these groups, even if the specificity does not correlate with any chemical properties of the R-groups. The 'blocky' structure, a known property of the genetic code, is composed of amino acids with utterly different chemical and structural properties of R-groups. Moreover, codons of one amino acid may occur in different blocks. Clearly, it is an artificial association of amino acid codons according to the principle of difference in at least one base out of three.

The above arguments favor the assumption that specificity of the genetic code is attributable to physical (van der Waals: dipole–dipole, induction, and dispersion) rather than chemical interactions. Amino acids share an important physical property of being dipoles at neutral pH. The magnitude of dipole moments and dipole oscillations depends on R-groups of amino acids; moreover, these parameters are different in the conformers of one and the same amino acid [24–26]. One can reasonably suggest that specific interactions of an amino acid (in the form of one of its conformers) with a given base sequence in the codon occurs only when the amino acid dipole oscillation frequency coincides with that of the induced dipole in the nucleotide triplet. This means that for the hypothesis of amino acid binding with codons to be verified, it is necessary to study the interaction between free L-amino acids and triplets of canonical nucleotides corresponding to their codons not only in terms of composition but also as regards nucleotide sequence. To the best of my knowledge, such studies have yet to be conducted.

For example, the authors of Ref. [27] undertook to elucidate the interaction between resin-immobilized amino acids and nucleotides and oligonucleotides. They revealed a binding of nucleotides and oligonucleotides with immobilized glycine and tryptophan that strengthened with increasing the number of nucleotides in the oligonucleotides. However, none of the three nucleotide triplets coincided with any codon of the two amino acids. In other words, the authors did not study the specificity of amino acid interaction with their codons. Immobilized amino acids are significantly different from free ones as far as conformers and dipole characteristics are concerned. This model appears to be at variance with real interactions between free amino acids and their codons.

Another line of research is represented by the work of Yarus and co-workers [28–30]. The first article of this series investigated the interaction between arginine and the GTP-binding site of ribosomal RNA [28]. It was revealed that arginine binds to this site via its R-group, whereas neither the amino nor carboxyl group is involved in the binding. Further studies brought to light RNA binding sites for seven more amino acids [30]. Although the authors emphasize that the binding sites are enriched in codons and anticodons for a given amino acid, other nucleotides participate in the binding, too; in other words, binding involves a huge (in comparison with the codon length) segment of the RNA molecule. Furthermore, such a binding depends on R-groups but not on amino and carboxyl groups. It may be conjectured that the binding of amino acids with RNA reported by Yarus et al. is an element of a regulatory mechanism, but it is doubtful that this type of binding is related to the origin of the genetic code.

To sum up, some research has demonstrated that amino acids interact with nucleotides, oligonucleotides, and RNA, but the interaction of amino acids with their codons awaits further studies.

One of the main factors responsible for the destruction of organic molecules in the prebiotic environment was UV radiation two orders of magnitude more intense than nowadays due to the absence of the protective ozone screen [31, 32]. Nitrogenous bases, ribose, and amino acids are known to absorb radiation in the UV region. The absorption of a quantum may either boost the molecule into the excited state (and even destroy it by breaking bonds) or result in dissipation of the absorbed energy into heat via molecule oscillations and rotation, the exchange of hydrogen atoms, and other mechanisms maintaining molecular integrity. The formation of molecular complexes (nucleotides, stacking of nucleotides, complementary nucleotide pairs) is associated with the increased number of interactions inside the complexes and, accordingly, enhanced dissipation of the radiation quantum energy into heat, accounting for a few orders of magnitude higher resistance of the complexes to UV radiation than that of separate molecules [33]. The binding of amino acids with codons could diminish the probability of their destruction and increase the lifetime of codons.

5. Some consequences of the bimolecular origin-of-life scenario

Let us consider one of the most important corollaries to the hypothesis in question. The hypothesis postulates that the reproduction of information occurred indirectly, i.e., was mediated through complementary codons. This means that the codons of all amino acids associated with life must have had complementary codons.

The table below includes pairs of amino acids having complementary codons. It illustrates the perfect correspondence between codons and complementary codons for 21 amino acids; there is not a single codon without a respective complementary codon. This fact is a strong argument in favor of the proposed hypothesis, because such a correspondence would have been impossible if amino acids had chosen codons at random. Moreover, the complete correspondence between codons and complementary codons for 21 amino acids suggests that all these amino acids were involved in the maintenance of life. The original coding was conserved during life evolution for 20 canonical, i.e., the commonest, amino acids.

With the advent of polypeptide synthesizing machinery in the course of evolution, it became necessary to arrest

Table. Amino acid pairs having complementary codons.

Amino acid	Codons	Complementary codons	Amino acid
Ala/A — alanine (nonpolar)[1]	**GCU**, **GCC**, **GCA**, **GCG**[2]	**CGA**, **CGG**, **CGU**, **CGC**, AGA, AGG	**Arg/R** — arginine (basic)
Ser/S — serine (polar)	**UCU**, **UCC**, *UCA*, *UCG*, *AGU*, *AGC*[3]	CGA, CGG, CGU, CGC, **AGA**, **AGG**	**Arg/R** — arginine (basic)
Asn/N — asparagine (polar)	**AAU**, **AAC**	**UUA**, **UUG**, CUU, CUC, CUA, CUG	**Leu/L** — leucine (nonpolar)
Asp/D — aspartic acid (acid)	**GAU**, **GAC**	UUA, UUG, CUU, CUC, **CUA**, **CUG**	**Leu/L** — leucine (nonpolar)
Glu/E — glutamic acid (acid)	**GAA**, **GAG**	UUA, UUG, **CUU**, **CUC**, CUA, CUG	**Leu/L** — leucine (nonpolar)
Cys/C — cysteine (polar)	**UGU**, **UGC**	ACU, ACC, **ACA**, **ACG**	**Thr/T** — threonine (polar)
Sec/U — selenocysteine (polar)	**UGA**	**ACU**, ACC, ACA, ACG	**Thr/T** — threonine (polar)
Trp/W — tryptophan (nonpolar)	**UGG**	ACU, **ACC**, ACA, ACG	**Thr/T** — threonine (polar)
Gln/Q — glutamine (polar)	**CAA**, **CAG**	**GUU**, **GUC**, GUA, GUG	**Val/V** — valine (nonpolar)
His/H — histidine (basic)	**CAU**, **CAC**	GUU, GUC, **GUA**, **GUG**	**Val/V** — valine (nonpolar)
Gly/G — glycine (nonpolar)	**GGU**, **GGC**, **GGA**, **GGG**, (UAA, UAG — green algae)	**CCA**, **CCG**, **CCU**, **CCC**	**Pro/P** — proline (nonpolar)
Gly/G — glycine (nonpolar)	GGU, GGC, GGA, GGG, (**UAA**, **UAG** — green algae)	AUA, **AUU**, **AUC**	**Ile/I** — isoleucine (nonpolar)
Ile/I — isoleucine (nonpolar)	**AUA**, AUU, AUC	**UAU**, UAC	**Tyr/Y** — tyrosine (polar)
Met/M — methionine (nonpolar)	**AUG**	UAU, **UAC**	**Tyr/Y** — tyrosine (polar)
Lys/K — lysine (basic)	**AAA**, **AAG**	**UUU**, **UUC**	**Phe/F** — phenylalanine (nonpolar)

[1] R-group.
[2] Complementary codons appear in boldface.
[3] Serine stabilizes not only complementary codons for arginine but also 4 mutually complementary codons (italicized), meaning that serine is the only amino acid capable of self-reproducing its information without intermediaries (other amino acids).

translation; this function passed to rarer codons, such as selenocysteine (UGA) and glycine (UAG) codons (conserved only in green algae), which became stop-codons. Selenocysteine underwent re-encoding by an additional nucleotide sequence (selenocysteine insertion sequence) following UGA. In such a combination, the protein synthesis apparatus interprets the UGA codon as coding a selenocysteine.

One amino acid (22nd, pyrrolysine) [34] appears to have been involved in the life process at a later stage during the evolution of one of the life tree branches, namely methane-producing archaea. The synthesis of pyrrolysine being a complicated process governed by several enzymes [35], it seems unlikely that it could be synthesized in a prebiotic medium. In this case, the UAG stop-codon had to be re-encoded to the pyrrolysine codon by the additional succeeding nucleotide sequence and thus enabled to be recognized as the archaean pyrrolysine codon. In other words, as new amino acids became involved in the life-forming process and underwent re-encoding under conditions of the already established protein synthesis, the presence of a base triplet alone was not enough: additional information (a nucleotide sequence) was needed.

It is currently believed that each amino acid is, as a rule, encoded by the first two nitrogenous bases, whereas the third one is of little consequence and can be substituted by mutations having no effect on the code. This accounts for the degeneracy of the genetic code. However, this rule holds only for codons of four amino acids: alanine, threonine, valine, and proline; it is not fulfilled in the case of any other codon. For example, not only the third but also the first base were replaced in arginine and leucine codons. The third bases in other amino acids were either less extensively substituted or remained unaltered (methionine, selenocysteine, tryptophan codons), despite the fact that all these codons evolved along similar pathways, which means that the difference in the number of codons per amino acid cannot be attributed to the substitutability of their third bases and the inability of the first and second ones to be replaced.

Let us assume that all the bases in all codons undergo substitution at a similar rate as a result of mistakes made by DNA and RNA polymerases and consider the simplest case, specifically, a single replacement of any of the three bases in the tryptophan codon, UGG (excluding two simultaneous replacements as unlikely). The substitution of all three bases gives rise to the following spectrum of codons: AGG (arginine), CGG (arginine), GGG (glycine), UAG (glycine), UCG (serine), UUG (leucine), UGU (cysteine), UGA (selenocysteine), and UGC (cysteine). Evidently, all newly formed (as a result of substitutions) codons are occupied by another amino acids; in other words, mutations result in the replacement of amino acids rather than the formation of a new (additional for a given amino acid) codon.

Here the questions arise: when codons became occupied by amino acids, and why the number of codons per amino acid is different? Suppose that degeneracy of the genetic code (an amino acid can be coded both by a single codon and by 2 and up to 6 codons) appeared at the earliest stages of life. This property of the genetic code is due to two factors: (1) different amino acids can stabilize codons by interacting not only with its three bases but also with two or three in various combinations, and (2) different conformers of an amino acid can stabilize different codons. The code is said to be nondegenerate when a single codon corresponds to three amino acids (methionine, tryptophan, and selenocysteine).

This means that to stabilize base triplets, these amino acids must interact with all three bases arranged in a proper sequence alignment (alternatively, only one of the conformers of these amino acids stabilizes the triplet of bases). Nine amino acids are encoded by two codons in which first two bases are identical and the third one is one of the two U–C or A–G pairs (this rule is followed in all nine codons). Clearly, two bases are insufficient to enable these amino acids to stabilize codons: a third base is needed. Any of the two pairs (or two conformers of each amino acid) are capable of stabilizing base triplets. Isoleucine is the sole amino acid having three codons; the first two bases are identical in all of them; the third base is represented by the U–C pair in two codons, and by A in the third one. For this amino acid, the third base can be any of the three (or three conformers of isoleucine stabilize base triplets). Finally, any third base suits well if interaction with only two is enough to stabilize the triplet. In this case, there must be four codons per amino acid: the first two being fixed, and the third any of the four (four conformers of the given amino acids stabilize base triplets). Alanine, valine, threonine, and proline have four codons each. Four amino acids are encoded by six codons. Leucine and arginine have two additional codons, besides the four, with the first bases U and C being replaced by C and A in the leucine and arginine codons, respectively. Two other amino acids having six codons (serine and glycine) deserve special consideration. Serine stabilizes not only the first two bases (UC) in its four codons, two of which are complementary to the two arginine codons, but also two codons complementary to its own two codons. It is the only amino acid capable of stabilizing complementary codons and thereby ensuring their replication without intermediaries. Glycine (except that found in green algae) has 4 codons sharing the first two bases (any other can be the third one). The first two glycine codons of green algae are quite different: their first two bases are UA instead of GG. It is the only known case when codons of one amino acid differ in the first two bases at one time. It cannot be a result of mutations following the split of green algae from other branches of the evolutionary tree, because each of the six glycine codons has a complementary codon; this means that glycine stabilized all six codons from the very beginning.

The assumption that replication was initially realized through the formation of complementary codons leads to the following conclusion:

(1) Arginine stabilizes 6 codons and reproduces 6 complementary codons, four of which are stabilized by alanine, and two by serine. Codons of these amino acids reproduce arginine codons (Arg, Ala, Ser: basic, nonpolar, polar).

(2) Leucine stabilizes 6 codons and reproduces 6 complementary codons, two of which are stabilized by asparagine, two by aspartic acid, and two by glutamic acid. Codons of these amino acids reproduce leucine codons (Leu, Asn, Asp, Glu: nonpolar, polar, acidic, polar).

(3) Threonine stabilizes 4 codons and reproduces 4 complementary codons, two of which are stabilized by cysteine, one by selenocysteine, and one by tryptophan (Thr, Cys, Sec, Trp: polar, polar, polar, nonpolar).

(4) Valine stabilizes 4 codons and reproduces 4 complementary codons, two of which are stabilized by glutamine, and two by histidine. Codons of these amino acids reproduce valine codons (Val, Gln, His: nonpolar, polar, basic).

(5) Glycine stabilizes 6 codons (two of them are conserved only in green alga) and reproduces 6 complementary codons, four of which are stabilized by proline, and two by isoleucine. Isoleucine, in turn, stabilizes 3 codons and reproduces one more complementary codon stabilized by tyrosine. The codon complementary to the tyrosine one is stabilized by methionine. Proline and isoleucine codons reproduce glycine codons, the tyrosine codon reproduces isoleucine codon, and the methionine codon reproduces the tyrosine codon (Gly, Pro, He, Tyr, Met: nonpolar, nonpolar, nonpolar, polar, nonpolar).

(6) Lysine stabilizes 2 codons and reproduces 2 complementary codons stabilized by phenylalanine. Phenylalanine codons reproduce lysine codons (Lys, Phe: basic, nonpolar).

(7) Serine stabilizes 4 mutually complementary codons and is therefore capable of self-reproduction (Ser: polar).

Thus, the proposition that there were nitrogenous bases and amino acids that were synthesized implies that life could emerge spontaneously in the form of seven SISs. The matrix (a triplet of nitrogenous bases) contained information about the structure of the operator (amino acid) reproducing this information. In six of these seven systems, information was reproduced by the operators indirectly through complementary codons and their amino acids. The appearance of SISs opened up the way for Darwinian evolution with competition between SISs for nitrogenous bases and amino acids, selection of information stability (survival), and reproduction rate (propagation), etc. The indirect reproduction of information being less feasible than the direct one, primary SISs most probably evolved via the merging of complementary codons into a unified complex. Another important factor that promoted the codon merging was UV radiation, because complexes are less sensitive to its destructive action than individual codons [30]. This gave rise to the emergence of RNA containing from 2 to 5 codons (or 6–15 nucleotides) and the respective new operators (peptides) to stabilize codon complexes. At this stage, self-reproduction of serine lost evolutionary significance, since it was a component of a tripeptide (system 1), whereas the future was behind codon complexes and peptides. Two scenarios for the creation of peptides are conceivable:

(1) Peptide bonds formed in interactions between amino acids and a codon complex during which the former were fixed and brought closer together, so that the amino group of one amino acid approached the carboxyl group of another (complexes with reverse orientation of codons with which amino acids came to be in contact through their amino or carboxyl groups were not reproduced). In this way, a correct orientation of codons in the matrix was achieved and has been retained up to now. The distance between stacked nitrogenous bases is 0.45 nm [21] or much greater than the peptide bond length (0.132 nm), but Coulomb interaction between oppositely charged amino acid groups could possibly bring neighboring amino acids close enough together for a peptide bond to form.

(2) Part of RNA resulting from codon merging was able to catalyze the formation of peptide bonds. It was shown in Refs [36, 37] that such catalytically active short-chain RNAs do exist.

The evolution of SISs toward direct reproduction gave rise to RNA and peptide molecules possessing structural and catalytic functions. However, the creation of an efficient metabolic ensemble was impossible without its isolation, because both catalytically active molecules and their products were dispersed by convection streams and currents. Therefore, the next step in the evolution of life should have

been its separation from the environment. Ever since, the atmospheric electricity and UV radiation have become the main sources of energy for abiogenic synthesis of organic compounds [38]. These reactions occurred in the atmosphere and resulted in the fallout of their products by precipitation over the surface of water bodies. Alternatively, they could proceed near the water reservoir surfaces, since UV radiation does not penetrate deep into water. It is known that fatty acids synthesized in abiogenic reactions tend to make up films on the surface of water. Some RNA and peptide molecules could enter bubbles formed when rain drops landed on water and were separated by such films from the liquid. Such closed space was not yet an operator, because the matrix contained no information about the film; initially, it was little more than an ecological niche. But even at this stage the film of fatty acids could undergo modification by peptides. It is noteworthy that all 1–6 systems generated amphiphilic peptides known to be able to build into fatty acid film and probably stabilize it by creating transport channels for the trafficking of amino acids and nucleotides.

6. Conclusion

The proposed hypothesis is the first version of the bimolecular origin-of-life scenario that is not, in the opinion of the author of Ref. [11], separated by a logical gulf from real life. According to this scenario, short-chain RNAs and peptides emerged simultaneously, and the subsequent merging of primary self-reproducing systems created a huge information potential for the evolution of life toward increasingly complicated operators, such as ribosomes, cells, plants, animals, and, finally, humans. Forestalling criticism of the hypothesis, it should be noted that its least convincing point is the assumed possibility of specific binding between L-amino acids and their codons (triplets of only canonical nucleotides: canonical bases with D-ribose), which allegedly increases the codon lifetime. Unfortunately, this question has not yet been addressed in either a theoretical or experimental context. It may be hoped that the proposed hypothesis will give an incentive for research in this field.

References

1. Ivanitskii G R *Phys. Usp.* **53** 327 (2010); *Usp. Fiz. Nauk* **180** 337 (2010)
2. Klyatskin V I *Phys. Usp.* **55** 1152 (2012); *Usp. Fiz. Nauk* **182** 1235 (2012)
3. Ivanitskii G R *Phys. Usp.* **55** 1155 (2012); *Usp. Fiz. Nauk* **182** 1238 (2012)
4. Jortner J *Phil. Trans. R. Soc. Lond. B* **361** 1877 (2006)
5. Eigen M *Naturwissenschaften* **58** 465 (1971)
6. Eigen M *Steps Towards Life*: *A Perspective on Evolution* (Oxford: Oxford Univ. Press, 1996)
7. Orgel L E *Trends Biochem. Sci.* **23** 491 (1998)
8. Lazcano A, Miller S L *Cell* **85** 793 (1996)
9. Ivanitskii G R *Virazhi Zakonomernostei*: *Pravilo BIO — Sterzhen' Nauki* (Overbanks of Regularities: The BHS Rule — the Pivot of Science) (Moscow: Nauka, 2011)
10. Gilbert W *Nature* **319** 618 (1986)
11. Cech T R *Gene* **135** 33 (1993)
12. Joyce G F *Nature* **418** 214 (2002)
13. Joyce G F, Orgel L E *The RNA World*: *The Nature of Modern RNA Suggests a Prebiotic RNA* (Eds R F Gesteland, T R Cech, J F Atkins) 2nd ed. (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press, 1999) p. 49
14. Walker S I, Davies P C W *J. R. Soc. Interface* **10** 20120869 (2013)
15. Saladino R et al. *Biochimie* **94** 1451 (2012)
16. Ponnamperuma C, Sagan C, Mariner R *Nature* **199** 222 (1963)
17. Powner M W, Gerland B, Sutherland J D *Nature* **459** 239 (2009)
18. Hein J E, Tse E, Blackmond D G *Nature Chem.* **3** 704 (2011)
19. Hein J E, Blackmond D G *Acc. Chem. Res.* **45** 2045 (2012)
20. Krishnamurthy R *Acc. Chem. Res.* **45** 2035 (2012)
21. Norberg J, Nilsson L *J. Am. Chem. Soc.* **117** 10832 (1995)
22. Woese C R *Proc. Natl. Acad. Sci. USA* **54** 1546 (1965)
23. Orgel L E *J. Mol. Biol.* **38** 381 (1968)
24. Gronert S, O'Hair R A J *J. Am. Chem. Soc.* **117** 2071 (1995)
25. Cao X, Fischer G *J. Phys. Chem. A* **103** 9995 (1999)
26. Kaczor A et al. *J. Phys. Chem. A* **110** 2360 (2006)
27. Saxinger C, Ponnamperuma C *Origins Life* **5** 189 (1974)
28. Yarus M *Science* **240** 1751 (1988)
29. Yarus M, Caporaso J G, Knight R *Annu. Rev. Biochem.* **74** 179 (2005)
30. Yarus M, Widmann J J, Knight R *J. Mol. Evol.* **69** 406 (2009)
31. Garcia-Pichel F *Orig. Life Evol. Biosphere* **28** 321 (1998)
32. Cockell C S, Horneck G *Photochem. Photobiol.* **73** 447 (2001)
33. Mulkidjanian A Y, Galperin M Y *Chem. Biodiversity* **4** 2003 (2007)
34. Atkins J F, Gesteland R *Science* **296** 1409 (2002)
35. Krzycki J *Curr. Opin. Microbiol.* **8** 706 (2005)
36. Turk R M, Chumachenko N V, Yarus M *Proc. Natl. Acad. Sci. USA* **107** 4585 (2010)
37. Yarus M *Phil. Trans. R. Soc. B* **366** 2902 (2011)
38. Chyba C, Sagan C *Nature* **355** 125 (1992)