Available online at www.sciencedirect.com
SCIENCE DIRECT®
Physics of Life Reviews 1 (2004) 202–229
PHYSICS of LIFE reviews
www.elsevier.com/locate/plrev

Genetic code: Lucky chance or fundamental law of nature?

Victor A. Gusev [a,*], Dirk Schulze-Makuch [b]

[a] *Sobolev's Institute of Mathematics of Siberian Division of Russian Academy of Sciences, Koptyuga pr. 4, Novosibirsk 630090, Russia*
[b] *Department of Geology, Washington State University, Webster Hall, Pullman, WA 99164, USA*

Abstract

The amount of publications devoted to the rise and structure of the genetic code is ever-increasing, which include semantic and structural analyses of the code as well as the problem of the origin of the code among others. The genetic code consisting of its triplet structure and canonical sets of nucleotides and amino acids was previously suggested to be a frozen accident or the result of accidental selection in the process of the evolution of a prebiotic system. These ideas are reviewed in this paper. It becomes clear that the information code is intrinsically related to the physical laws of the universe, and thus life may be an inevitable outcome of our universe. The lack of success in explaining the origin of the code and life itself in the last several decades suggest that we miss something very fundamental about life, possibly something fundamental about matter and the universe itself. Certainly, the advent of the genetic code was no "play of chance".

Keywords: Genetic code; Evolution; Life; Nucleotide; Information

> The problem of the genetic code has several facets, of which the most compelling is why it is just what it is.
> J. Maddox

* Corresponding author.
E-mail address: vgus@math.nsc.ru (V.A. Gusev).

1571-0645/$ – see front matter © 2004 Published by Elsevier B.V.
doi:10.1016/j.plrev.2004.11.001

1. Introduction—formulation of the problem

"*The origin of the code is perhaps the most perplexing problem in evolutionary biology. The existing translation machinery is at the same time so complex, so universal, and so essential that is the hard to see how it could have come into existence, or how life could have existed without it*" [93].

The structure of the genetic code was completely deciphered in the late 1960s and the table of the correspondence between triplets and amino acids attained a canonical form as shown in Fig. 1.

Further efforts of researchers were directed to explain observed regularities. Three well-known hypotheses were advanced to understand the essence of correlation between codons and amino acids. They are the (1) stereo-chemical hypothesis; (2) frozen-accident hypothesis or hypothesis of accidental freezing; (3) evolutionary hypothesis [2,3,6,13,20,24,25,36,92,100,103]. Some of these theories considered an evolution of the gene code based on relationships between codons and amino acids, dependencies of amino acid biosynthesis and the ways how metabolic pathways could be selected by evolution, or postulating a frozen-accident relationship between amino acids and codons.

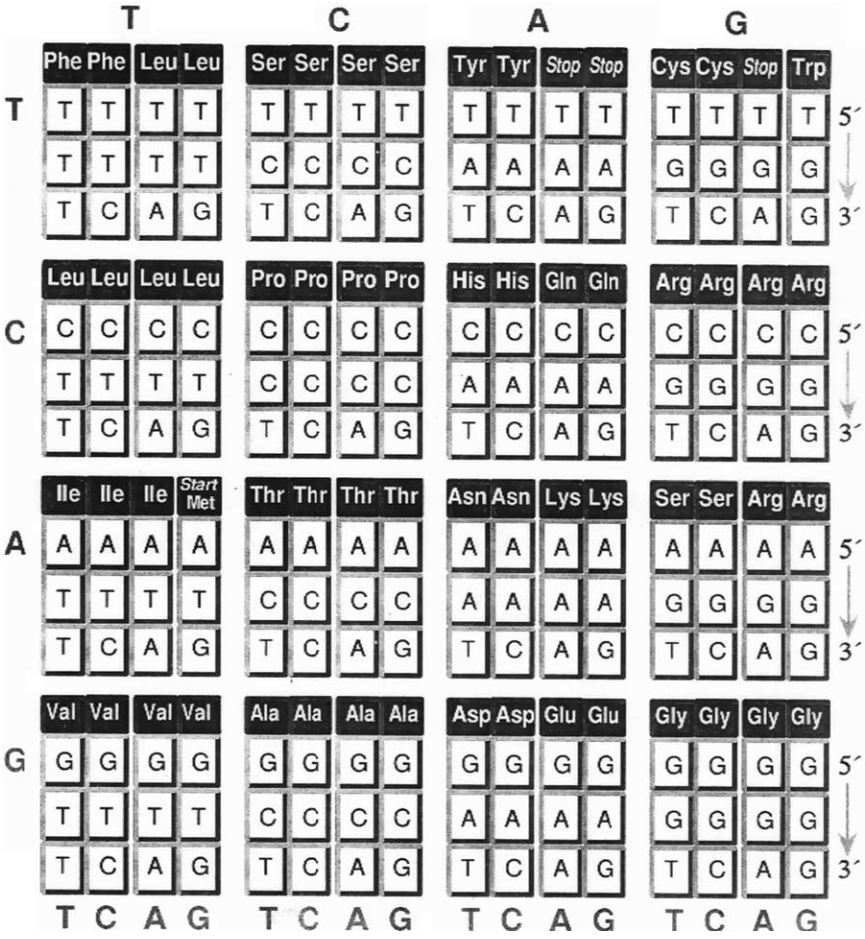

Fig. 1. Table of canonical genetic code.

A priori, all hypotheses postulated the triplet structure of codons. As a background of this postulate, serves the "almost evident" assumption made by George Gamov [39] after Watson and Crick have published their famous article about the DNA structure [97]: for a one-lettered code, the number of combinations of 4 by 2 equals four; whereas for the doublet code, this number equals 16. For a triplet code the total number of variations is 64, which is more than sufficient for coding 20 amino acids. If a quadruplet code is used, this would result in excessive redundancy.

A trivial phrase by J. Maddox, placed as an epigraph [63] can be extended in a form of "simple and evident" questions, to which contemporary science on living matter does not know any consequent and self-consistent answer:

1. Why is the genetic code universal?
2. Did the dialects, i.e., mitochondrial version, with UGA codon (being the stop codon in the universal version) codifying tryptophan; AUA codon (being the isoleucine in the universal version), methionine; and *Candida cylindrica* (funges), with CUG codon (being the leucine in the universal version) codifying serine, appear accidentally or as a result of some kind of selection process?
3. Why is the genetic code represented by the four bases A, T(U), G, and C?
4. Why does the genetic code have a triplet structure?
5. Why is the genetic code not overlapping, that is, why does the translation apparatus of a cell, which transcribes information, have a discrete equaling to three, but not to one?
6. Why does the degeneracy number of the code vary from one to six for various amino acids?
7. Is the existing distribution of codon degeneracy for particular amino acids accidental or some kind of selection process?
8. Why were only 20 canonical amino acids selected for the protein synthesis?
9. Is this very choice of amino acids accidental or some kind of selection process?

In the review presented here, we have made an attempt to provide possible solutions and avenues to at least in part, address these questions related to the origin of life on Earth and its essential attribute, the genetic code.

2. The triplet code structure of the four-letter code

Arzamastsev [10], Aldana et al. [4], and Aldana-Gonzales et al. [5] have made an attempt to place the postulate by G. Gamov into the rank of interpreted categories, that is, to find the physical reasoning explaining why the triplet code composed out of four symbols is optimal. The basic idea by Arzamastsev [10] resides in looking for a compromise between low number of letters in an alphabet (n), which simplifies it as a decoding machine, but generates lengthy informational sequences, and a large n, which decreases the sequence length, but complicates the informational machine. By admitting the postulate that protogenomes consisted of relatively short sequences, of three-four thousands letters, the author demonstrates that the optimum indicated is achieved, if the code is four-lettered.

Arzamastsev [10] operates within the limits of the frozen-accident hypothesis, because in his model he optimized informational capacity of symbolic sequence and decoding pattern. Quoting him: "*the situation when Nature invented the DNA code surprisingly resembles designing a computer by man. If a computer were designed today, the binary notation would be hardly used. Binary notation was chosen only at the*

first stage, for the purpose to simplify at most the construction of decoding machine. But now, it is too late to correct this mistake". This viewpoint could be flawed, however, since it has a serious shortcoming. Based on Arzamastsev [10] the primary genomes that originated on Earth were those of viruses. However, the virus genome, as it is known, is not all-sufficient for the autonomous reproduction outside the host cell. Although, this defect is typical for all evolutionary hypotheses of life's origin based on the principle "from simple towards complicate".

In order to explain the triplet structure of the four-lettered code Aldana et al. [4] and Aldana-Gonzales et al. [5] proposed the idea of dynamical or powered communication between electrophysical properties of the matrix polymer and the decoding molecular device. The essence of the model is to measure an electrostatic interaction between an oligomer consisting of M monomers (M varies within the range from 1 to 10–20) and long polymer molecules ($M \gg 1$), compiled from the same monomers. The interaction potential V between the ith monomer of the oligomer and jth monomer of the matrix is expressed by the following equation:

$$V_{ij}(x) = p_i q_j \int_{x+i-1}^{x+i} \int_{j-1}^{j} \frac{\mathrm{d}x'\,\mathrm{d}x''}{[(x'-x'')^2+\sigma^2]^{\alpha/2}}, \tag{1}$$

where p_i and p_j are the values of the monomer's electrostatic charge, x is a linear coordinate across the matrix, and σ is the distance between oligomer and polymer matrixes. The parameter α characterizes the kind of electrostatic interaction between these chains: $\alpha = 1$ corresponds to an ion–ion interaction, $\alpha = 2$ represents an ion–dipole interaction, and so forth. The total potential of interaction between the matrix and oligomer is determined by the sum of interactions between each monomer consisting of the oligomer and every monomer of the matrix. Numerical modeling is based on the distribution of potential energy minimums over a polymer matrix, in case oligomers of different lengths slide through the matrix. As matrix a sequence of oligomers was used, which would be arbitrary, in principle. Despite the simplicity and demonstrativeness of the model, the authors have obtained rather intriguing results. If the matrix consisting of four different monomers is distributed stochastically along the polymer, the probability distribution function of two consecutive minima is maximal if the monomers are separated by three others. Following Newton's equation, Aldana et al. [4] and Aldana-Gonzales et al. [5] have shown that for the oligomer's moving along the matrix, the time spent by an oligomer at the potential minimum is by 2–3 times greater than in the intermediate condition. The authors suggest that this fact may serve as a precondition for the triplet organization of the genetic code.

Analysis of the probability function of a potential energy minima distribution of a real genome, for example *Drosophila melanogaster*, has revealed the difference between different DNA regions, exons and introns [5]. In the protein coding regions, the maxima of function correspond to the triplet interchange, whereas for the non-coding DNA, this distance is less pronounced. In the protein coding regions of *Drosophila melanogaster*, *Chlorella vulgaris*, *Deinococcus radiodurans*, *Mycobacterium tuberculosis*, and *Aeropyrum pernix* genomes, the peak of the probability function for potential energy minima was four times as high as the value for the random nucleotide sequences [5].

The data obtained in the studies by Aldana et al. [4] and Aldana-Gonzales et al. [5] should be interpreted that the three-base codon structure of the genetic code is determined by the physics of the interaction between monomers at the early stages of the heteropolymer matrix synthesis. In other words, selection of codon structure on Earth took place during the chemical, but not the biological evolution. Also, the difference in probability function distribution for the distances between the nearest energy

minima calculated for the random sequences, coding and non-coding genome regions makes us to suppose that the distribution pattern of monomers in the polymer matrix is determined. Thus, an arbitrary sequence of four monomers, which are formed by triplets, could be translated by a hypothetical enzyme apparatus like a ribosome, in principle. As a result, a polypeptide chain lacking any functional role could be obtained. Only in case the spatial protein core and the active protein center will be formed, the polypeptide will function as an enzyme. These events could occur if amino acids in a polypeptide will occupy specific spatial positions. Since the distribution function of triplet alternation of potential energy minima occupies a peak exactly at the sensing sequence of monomers, we may suppose that the structure of functionally important proteins (hence, the functionally significant sequences of monomers!) is determined, at least in part, by the physics of interaction between the monomers.

3. Amino acids and polypeptides

An idea of antecedence of polypeptide spatial structure for genetic code semantic structure was formulated in the studies made by several Japanese scientists [43–46]. As the basic parameters determining spatial protein structure, they analyzed six parameters of proteins: hydrophilicity/hydrophobicity, α-helix, β-sheet, β-turn, and acidic or basic amino acid contents. Based on their analysis, these parameters determine exactly the folding and properties of a native polypeptide chain.

By comparing these native protein parameters in *bacteria* and *archaea* with GC-pair content of encoding genes (all in all, seven genomes were analyzed), they concluded that the parameters indicated no change, if the GC-pair content varies within the range of 20–75%. Simultaneous estimation of stop-codon occurrence at the non-coding chain revealed that if the GC content is above 60%, then the stop-codon could be met in DNA of 300 bp in length with a negligibly small probability.

Based on these data, Ikehara et al. [45] and Ikehara [43] propose several hypotheses: (1) *Novel genes in contemporary microorganisms may evolve from the non-coding DNA chain*; (2) *At early stages of prebiological evolution, primary genes could be formed through GC-rich polymer matrices with relatively monotonous structure, like SNS, where S corresponds either to G or to C, and N position is occupied by any out of four bases*. This codon structure was suggested after comparing base positions in the triplets of GC-rich genomes of bacteria and archaea. As found, in the first and third positions, GC-pair content is 5–10 fold higher than AT content. At the second position, all the bases occurred with the same frequency.

SNS structure of the universal genetic code precursor is represented by 16 codons encoding 10 amino acids: Leu, Pro, Val, Ala, His, Gln, Asp, Glu, Arg, and Gly. Based on Ikehara [44], Ikehara et al. [43], Ikehara et al. [45], and Ikehara [46] this set of amino acids possesses the properties that are necessary to form the tertiary structure of water-soluble polypeptides. Further evolution of the SNS precursor towards the universal genetic code NNN may occur in the process of GC-pair content drawdown (or, following the authors' expression, due to "*AT-mutation pressure*") under multiple reproduction of primary nucleotide sequences.

The ideas described in [43–46] are very original and rather constructive. The disadvantage of this approach is that it a priori assumes the triplet structure of the genetic code. However, if we compare the results of these works with conclusions made in Aldana et al. [4] and Aldana-Gonzales et al. [5] about a possible determination of the code's structure by the physics of the interaction between the primitive translating molecular apparatus and the polymer matrix, then this assumption does seem reasonable. From this viewpoint, the above studies made by two groups of authors supplement each other.

The SNS-NNN hypothesis [43–46] is constructed following the commonly accepted evolutionary principle of *from simple to complicated or from general to special*. Since the first live forms appeared on the Earth, further progressive evolution of biota followed exactly this principle. However, that does not necessarily mean that the formation of primary organelles, and in particular, the carriers of genetic information, followed the same scenario. Also, the environment under which the genetic code and life in general originated may have been very different from the Earth's environment today, and may in fact not exist anymore.

4. The degeneracy of the genetic code

In a series of articles by Ardell and co-authors (e.g., [7–9,83]), the degeneracy of the modern genetic code is considered as a result of co-evolution of genetic information contained in the primary nuclear sequences and an increase in its volume due to "dropping down" of codon degeneracy. As supposed a priori, the primary cellular organelles responsible for translation of genetic information were already existing.[1] These publications followed Crick's hypothesis claiming that the primary code was less degenerate than its contemporary version [20]. During the process of biological evolution, the structural and functional variability of proteins was growing due to an increase in number of amino acids used. In its turn, this led to an increase in genetic code degeneracy up to its present condition. Next, according to Crick [20], a phenomenon of "*accidental freezing of codon–amino acid correspondence*" took place. However, this conclusion may only be a philosophical allegory based on intuition. As a counter-argument to this statement, Ardell [7], Ardell and Sella (2001), Sella and Ardell (2002), and Ardell and Sella (2002) claimed that analytical calculations prove that the modern version with its degeneracy degree is a result of the evolution of the primary code via attenuation and by optimization of mistakes occurring both during genome translation and transcription.

In fact, a mathematical model of genetic code optimization [7–9,83] points to a class of informational models, because the problem of minimization of the noise effect under multiple transmission of equitypical information is solved. As an informational source, a nucleotide sequence of an arbitrary length (by definition of the authors a "message") is considered. Each codon with given degeneracy corresponds to a particular amino acid. Then a multiple reproduction of this sequence takes place against the background of mutations in form of nucleotide transitions. The extent of degeneracy is variable under various environmental condition. Under natural conditions alterations in the *message*, or amino acid sequence of the protein, are the quantitative optimization criterion. The authors apply physicochemical homology of amino acids, expressed as polarity of side-chain groups. They consider permissible amino acid substitutions with closely matched physicochemical properties. One of their main conclusions is that a triplet–amino acid correspondence characteristic for the canonical genetic code could most likely not be explained by stereochemical correspondence between amino acids, but by optimizing protection of genetic information from mutational noise under its repeated transmission.

This series of works [7–9,83] appears convincing from the viewpoint of a possible way of evolving a primary genetic code to its optimal condition, but it tells nothing about the initial state of the code that could be transformed to its present state. However, Ardell and Sella [8] point to further work in

[1] Notably, this formulation of a problem could not answer the question on the evolution of the translation machine. Thus, the problem is equivalent to the problem, which was first—chicken or egg?

this respect: "*Further work could examine the extent to which the size and dimensionality of codon and amino-acid spaces, the form of fitness interactions of amino acids within and among sites, and the particulars of the code-message coevolutionary dynamics affect the results and hence the generality of our interpretations*".

5. Coevolution and the question of the optimum code

The principle of coevolution explaining the origin of the modern code table is advanced in Di Giulio and Medugno [26,27], Di Giulio [28,29]. According to them, coevolution took place not at the level of the structure as proposed in Ardell [7], Ardell and Sella [8,9], Sella and Ardell [83], but at the level of biochemistry. The authors are supported by the studies from Wong [103,105] and Taylor and Coates [96], where five families of amino acid synthesis pathway were analyzed: (1) Phe-Tyr; (2) Ala-Val-Leu; (3) Ser-Gly-Cys-Trp; (4) Glu-Gln-His-Pro-Arg; (5) Asp-Asn-Lys-Thr-Ile-Met. Each biochemical amino acid family was associated with its own group of synonymous codes. By analogy to the previously described series of publications, here the authors compare amino acids' physicochemical properties, i.e., polarity of side residues and their spatial dimensions. By varying codons through altering nucleotide content and, hence, amino acids corresponding to these nucleotides, the authors minimize the function to the form:

$$\Delta(p) = \sum_{\substack{i,j=1 \\ i<j}}^{20} a_{ij}\left(|P_{p(i)} - P_{p(j)}|/\sigma_p + |V_{p(i)} - V_{p(j)}|/\sigma_v\right) \Bigg/ \sum_{\substack{i,j=1 \\ i<j}}^{20} a_{ij}, \tag{2}$$

where p is the number of permutations on the indices i and j of P and V; P_i, P_j are the polarity values [101] of the ith and jth amino acids, respectively; V_i, V_j, are molecular volumes [40] of the ith and jth amino acids; σ_p, σ_v, are standard deviations; and a_{ij}, is the number of times that amino acid i transforms into amino acid j when the codons of amino acid i transform into those of j through single-base changes. As shown in Eq. (2), the minimal value of the function is attained if physicochemical properties of amino acids, characterized by polarity and molecular volume, are most similar. Wong [103,105] and Taylor and Coates [96] issue the challenge whether it is possible to prove that the standard genetic code is optimal. Is the genetic code a local minimum? And is it the absolute minimum obtainable under certain conditions starting from any code configuration? If we suppose that in the process of the coevolution of the code, the "adjustment" of codons and corresponding to them amino acids was produced at the level of biochemical pathways, then, based on the existence of five biochemical families, the total number of possible codes is estimated as $6!5!4!3!2! = 24883200$. Such enormous variability of potentially possible codes prevents us to consider that the modern form of the code is completely optimal. In this sense, the frozen-accident hypothesis looks like a quite convincing supposition.

Freeland and his co-workers took recently a different approach to test the efficiency and total number of possible genetic codes [36–38,53]. Their approach was computer-based with the basic objective to randomize the genetic code and then compare the efficiency of a certain fraction of the vast number of alternatives codes the computer can generate to the natural code for life on Earth. The implicit assumption was that the genetic alphabet is composed of two base pairs (AT/CG; [94,95]), as well as the system of triplet codons and the 20 amino acids (e.g., [99]) available for protein construction found in all Earth life to be representing some sort of norm. Freeland and Hurst [37] examined a million alternative codes. At

first approximation the resulting distribution could be compared to a familiar normal distribution. However, the result was quite startling. They wrote "*the natural genetic code shows evidence of optimization. Two orders of magnitude higher than has been suggested previously. Through the precise quantification used here may be questioned, the overall result seems fairly clear: under our model, of* 1 *million random variant codes produced, only* 1 *was better . . . than the natural code—our genetic code is quite literally* 1 *in a million*" [37]. They further concluded that "*not only that the natural genetic code is extremely efficient at minimizing the effects of errors, but also that its structure reflects biases in these errors, as might be expected were the code the product of selection*". It appears a priori extremely unlikely that any two biospheres—separated by a gulf of many light years—would arrive at the same evolutionary solution, but it would be even more improbable that the solution achieved was not only good but in fact the very best [19]. Yet, the work by Freeland and co-workers may point toward this direction. On the other hand, the one code that was identified and that might in principle be better than the natural code has little similarity with the one used by life on Earth. Also, the 1 million alternatives tested by Freeland and Hurst [37] is only a small subset of all possible codes. When completely randomizing the existing code, it suggested a total number of about 10^{18} possible codes. However, in a subsequent analysis they suggested that the number of alternative codes that overall are realistically functional is relatively small, about 270 million.

Statistical analysis of possible codes by Di Giulio and Medugno [26,27], and Di Giulio [28,29] are consistant with the findings by Freeland and co-authors [38] that no unambiguous conclusions about the uniqueness of the genetic code can be drawn. In their work they do, as discussed above, attempt to transform the coevolutionary postulate into the rank of provable categories, that is, to validate numerically that existing correspondence between amino acids and codon encoding is a result of biochemical selection, but not stereochemical conformity. Di Giulio and Medugno [27] formulate it as follows "*The coevolution theory* [103] *predicted that the physicochemical properties of amino acids must be linced to the organization of the genetic code, although this correlation must not exceed a certain level* [104]. *If this were to happen, the stereochemical* [13,92,102,107] *and physicochemical* [35,50,55,56,98,100,101] *theories would be better supported. As we have noted no particular behavior in the physicochemical properties of the amino acids in the genetic code table for the smallest definable set of codes, we are led to conclude that the link between amino acid properties and genetic code organization is better explained by the coevolution theory* [103] *then by the physicochemical postulates* [13,35,50,55,56,92,98,100–102, 107]. *In other words, it is in the very set of codes subject to biosynthetic constraints that the properties of amino acids must have shown an extremely strong link to genetic code organization because this set is the smallest finable one and it has also almost certainly had an extremely important evolutionary significance. Nevertheless, the minimization percentage and much of the analysis performed in this paper are not such as to corroborate the conclusion that the allocation of amino acids in the genetic code was fundamentally determined by their physicochemical properties, for instance, via a selective pressure aiming to minimize the deleterious effects of translation errors. We, therefore, continue to favor the postulates of the coevolution theory*" [23,103].

This rather extensive quotation is given here for demonstrating that the general arguments settled for adopting this or that hypothesis on the origin and evolution of genetic code are strongly based on personnel sympathies of authors, but not on strict scientific reasoning.

6. The fundamental problem of the origin of the code

Thus, based on the discussion above, neither of the commonly adopted hypotheses, i.e., stereochemical hypothesis, accident freezing hypothesis, or evolutionary hypothesis, is able to answer the rather obvious 9 questions formulated at the beginning of this article (please see Introduction). Each hypothesis enables to solve a local problem, but all in all, the problem of the origin of the genetic code stays terra incognita.

Notably, all these hypotheses propose explicitly or implicitly the pre-existence of a molecular apparatus of self-reproduction, that is, transcription and translation of some primitive objects were needed for evolving the genetic code. This veil of silence, "or sweeping the litter under a carpet", during fifty years after discovery of the DNA structure and the publication of the idea on triplet organization of the genetic code [39,97] appears to indicate that by applying biological categories alone, the problem of the origin of the genetic code, and with it the problem of life's origin may be unsolvable.

As an illustration, let us consider the historical example how the basics of quantum mechanics were formed. By the end of 19th century, a "critical mass" of optical experimental data accumulated that could not be interpreted by laws and principles of classic physics. To explain these data adequately, Max Plank worked on the formulation of an absolutely new fundamental hypothesis of quanta. It took over twenty years to pass from working hypothesis to theory. In our case here, the period without a breakthrough is lasting already for half of a century without a solution in sight, hence, clearly pointing out that we are missing something very fundamental how the genetic code and life itself originated.

7. Life from a physical viewpoint

Life from the viewpoint of physics was first considered by Schrödinger [80]. Later, the fundamental idea of filliping group theory analysis of the genetic code structure was formulated by Rumer[2] [78]. He was the first to determine the operation of purine-pyrimidine inversion C ↔ A, G ↔ T(U) (now known as "Rumer's transformation"), which unambiguously transforms one into another two groups of octets from the genetic code table. The theoretical-group analysis consists of searching for symmetries and numerical regulations in the structure of the genetic code given in the matrix representation [31,32,48,49,68–71]. Any symmetry in a structure or a process points to the fact that the laws exist for selection or particular prohibitions [33]. In other words, the symmetry is evidence for some conservation law, however, the physical meaning of this law is not always obvious. In the works by Duplij and Duplij [31,32], the authors introduce a vector representation of four codons in a form of a column vector

$$\vec{V} = \begin{pmatrix} \mathrm{C} \\ \mathrm{G} \\ \mathrm{U} \\ \mathrm{A} \end{pmatrix}$$

[2] In [22], Danckwerts and Neubert independently re-opened the purine-pyrimidine symmetry of genetic code.

and, respectively, a row vector $\vec{V}^T = (\mathrm{C\ G\ U\ A})$.[3] The exterior product of these vectors is given by the matrix of doublets:

$$\begin{pmatrix} \mathrm{CC} & \mathrm{CG} & \mathrm{CU} & \mathrm{CA} \\ \mathrm{GC} & \mathrm{GG} & \mathrm{GU} & \mathrm{GA} \\ \mathrm{UC} & \mathrm{UG} & \mathrm{UU} & \mathrm{UA} \\ \mathrm{AC} & \mathrm{AG} & \mathrm{AU} & \mathrm{AA} \end{pmatrix}. \tag{3}$$

If each base of a vector is characterized by a number [31,32], or if in accordance with Rumer [79], we introduce the power of a base, d, given in a form $d_\mathrm{C} = 4$, $d_\mathrm{G} = 3$, $d_\mathrm{A} = 2$, $d_\mathrm{U} = 1$, whereas the power of a doublet is determined as the sum $d_{ij} = d_i + d_j$, where $i, j \to \mathrm{A, U, G, C}$, then the matrix (3) could be represented in a digital form:

$$\begin{pmatrix} 8 & 7 & 6 & 5 \\ 7 & 6 & 5 & 4 \\ 6 & 5 & 4 & 3 \\ 5 & 4 & 3 & 2 \end{pmatrix}. \tag{4}$$

Let us note the main properties of this matrix: "*the matrix symmetry is so high that the matrix is singular, that is, its determinant equals to zero,* $\det D = 0$; *matrix rank equals to* 2, $\operatorname{rank} D = 2$; *and excess of this matrix also equals to* 2. *This is a consequence of the fact that the matrix is an exterior product of the vectors* (3). *Noteworthy, the spur of the matrix equals to* $\operatorname{tr} D = 20$ *and it coincides to the sum of the elements from the matrix side diagonal. As seen, by the side diagonal of the matrix* (4), *as well as in parallel to it, the doublets are* "*equal in power*". *Matrix transposition relatively the side diagonal corresponding to the purine–pyrimidine inversion does not change the main matrix properties*" [31].

Obviously, the operation of exterior production of matrix (3) to one of the vectors will represent the three-dimensional matrix consisting of all 64 triplets. In fact, the authors introduce their own version of a graphical representation of the genetic code table in form of a three-dimensional cube, such that its edge contains 16 triplets. The complete list of symmetries and numerical regularities of particular elements of the cubical version can be found in the original literature by Duplij and Duplij [31,32] and is not presented here. Below, we cite only their conclusions that *the algebraic approach enables to give us another look on the problem of the genetic code*: "***Symmetries observed in the code reveal them in the process of choosing codons for determination of various amino acids***. *The model suggested is designed on a different ability of a nucleotide,* C, G, U, *or* A, *to determine unambiguously an amino acid. The introduced notion of determinative degree of a nucleotide enables to represent genetic texts as the consequence of numbers ranging from* 1 *to* 4 (*actually, in the quaternary number system*). *Analysis of such sequences may lead to better understanding of transcription and, possibly, to formulation of novel principles of constructing recombinant DNA, an essential component of developing modern methods of cloning and gene engineering*" [31].

The author's most profound statement, marked by us in bold, can be viewed as an argument that the genetic code is not a frozen accident, but on the contrary, the existing correlations between number of codons and respective amino acids are probably a result of some selection process. Still, the possibility remains to explain the essence of this phenomenon by means of biological or physical notions.

[3] Under such vector definition, an analogy to Dirak's matrix representation in quantum mechanics is clearly seen.

8. The group theory approach

In a following series of publications Negadi [68–71] applies the group-theory approach for studying the structure of the genetic code. He analyzes not only the code itself, but the chemical structure of nitrogenous bases as well. His objective was to find logically, but not semi-empirically, the relationships between the chemical content of nitrogenous bases and the genetic code structure.

Negadi [68–71] applies matrix representation for each chemical element composing nitrogenous base in the following way:

$$\mathrm{H} \equiv \begin{pmatrix} 1(\mathrm{H}) & 0 \\ 0 & 0 \end{pmatrix}, \qquad \mathrm{C} \equiv \begin{pmatrix} 0 & 6(\mathrm{C}) \\ 0 & 0 \end{pmatrix}, \qquad \mathrm{O} \equiv \begin{pmatrix} 0 & 8(\mathrm{O}) \\ 0 & 0 \end{pmatrix}, \qquad \mathrm{N} \equiv \begin{pmatrix} 7(\mathrm{N}) & 0 \\ 0 & 0 \end{pmatrix}, \tag{5}$$

where the values near the symbol of the atom denote its number in Mendeleev's table. The "dual" matrices (5) are determined, which are calculated by two-side multiplication to the conjugation matrix

$$\eta = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{6}$$

An example of the matrix that is dual to the matrix of hydrogen looks as follows:

$$\tilde{\mathrm{H}} \equiv \eta \mathrm{H} \eta^{-1} = \begin{pmatrix} 0 & 0 \\ 0 & 1(\mathrm{H}) \end{pmatrix}. \tag{7}$$

The remaining dual matrices can be obtained by analogous multiplication. Any chemical compound may be represented in a matrix form. Below, one may see the examples of matrix representation of the nitrogenous bases:

$$\begin{aligned}
&\mathrm{U} = (\mathrm{C}\tilde{\mathrm{C}})^2\mathrm{N}^2\mathrm{H}^4\mathrm{O}\tilde{\mathrm{O}} \equiv \begin{pmatrix} \mathrm{U} & 0 \\ 0 & 0 \end{pmatrix}, \qquad \tilde{\mathrm{U}} = (\tilde{\mathrm{C}}\mathrm{C})^2\tilde{\mathrm{N}}^2\tilde{\mathrm{H}}^4\tilde{\mathrm{O}}\mathrm{O} \equiv \begin{pmatrix} 0 & 0 \\ 0 & \mathrm{U} \end{pmatrix}, \\
&\mathrm{C} = (\mathrm{C}\tilde{\mathrm{C}})^2\mathrm{N}^3\mathrm{H}^5\mathrm{O} \equiv \begin{pmatrix} 0 & \mathrm{C} \\ 0 & 0 \end{pmatrix}, \qquad \tilde{\mathrm{C}} = (\tilde{\mathrm{C}}\mathrm{C})^2\tilde{\mathrm{N}}^3\tilde{\mathrm{H}}^5\tilde{\mathrm{O}} \equiv \begin{pmatrix} 0 & 0 \\ \mathrm{C} & 0 \end{pmatrix}, \\
&\mathrm{A} = (\mathrm{C}\tilde{\mathrm{C}})^2\mathrm{C}\tilde{\mathrm{N}}^5\tilde{\mathrm{H}}^5 \equiv \begin{pmatrix} 0 & \mathrm{A} \\ 0 & 0 \end{pmatrix}, \qquad \tilde{\mathrm{C}} = (\tilde{\mathrm{C}}\mathrm{C})^2\mathrm{N}^5\mathrm{H}^5 \equiv \begin{pmatrix} 0 & 0 \\ \mathrm{A} & 0 \end{pmatrix}, \\
&\mathrm{G} = (\mathrm{C}\tilde{\mathrm{C}})^2\mathrm{C}\tilde{\mathrm{N}}^5\tilde{\mathrm{H}}^5\tilde{\mathrm{O}} \equiv \begin{pmatrix} \mathrm{G} & 0 \\ 0 & 0 \end{pmatrix}, \qquad \tilde{\mathrm{G}} = (\tilde{\mathrm{C}}\mathrm{C})^2\tilde{\mathrm{C}}\mathrm{N}^5\mathrm{H}^5\mathrm{O} \equiv \begin{pmatrix} 0 & 0 \\ 0 & \mathrm{G} \end{pmatrix}.
\end{aligned} \tag{8}$$

By summing up corresponding matrices, two base matrices were obtained:

$$\mathrm{B} \equiv \mathrm{U} + \mathrm{C} + \tilde{\mathrm{A}} + \tilde{\mathrm{G}} = \begin{pmatrix} \mathrm{UC} \\ \mathrm{AG} \end{pmatrix}, \qquad \tilde{\mathrm{B}} \equiv \tilde{\mathrm{U}} + \tilde{\mathrm{C}} + \mathrm{A} + \mathrm{G} = \begin{pmatrix} \mathrm{GA} \\ \mathrm{CU} \end{pmatrix}. \tag{9}$$

The matrix form given above represents four bases and illustrates base matrices in numerical form without introducing the somewhat undetermined definition of the base power [31,32]. For example, as follows from (8),

$$\mathrm{C} = (\mathrm{C}\tilde{\mathrm{C}})^2\mathrm{N}^3\mathrm{H}^5\mathrm{O} = 6^4 7^3 8 = 3556224. \tag{10}$$

Then the base matrix may be represented in a form:

$$B = \begin{pmatrix} UC \\ AG \end{pmatrix} = \begin{pmatrix} 4064256 & 3556224 \\ 130691232 & 1045529856 \end{pmatrix}. \quad (11)$$

The meaning of these long numerical series and their possible combinations are extensive enough to be the material of a separate paper. However, representation of the genetic code in a matrix number system anchors the hopes that in future the relationships will be found between positioning in Mendeleev's table (i.e., factually, the charge of an atom) of the basic chemical elements C, N, O, and H, which compose the nitrogenous bases, and the structure of the code itself.

It should be noted that remarkable "magic numbers" were detected by Negadi [68–71] in the course of analyzing the genetic code structure by the group theory approach.[4] "Magic number" is a recognized term in physics. "Magic numbers" characterize the numbers of discrete units, i.e., electrons in an atom or nucleons in a nucleus, which occupy the same quantum energy level. In other words, these numbers characterize the stable state of elementary particles of a substance. Negadi went on and compared the "magic numbers" found by him with the analogous numbers in physical models, in particular, supergravity theory. The numbers 44, 128, and 84 characterize the degrees of freedom fora graviton, gravitino, and gauge field in supergravity theory, respectively [69]. In the model by Negadi [69] the same numbers characterize complete sums of matrix elements of the patterns, which could be transformed into each other by Rumer's transformation. The author views the magic numbers emerging in a group theory analysis of the genetic code and their coincidence, in some cases, with the numbers occurring in physical models and describing the structure of creation, as "curious incidents", though... "*Now, concerning the interpretation*(*s*) *of this new kind of small numbers coincidences, we have proposed some personal thought coming from a theoretical physicist which could be, perhaps, one of the theoretical physicists undergoing the great migration towards Biology, Erwin Schrödinger announced a long time ago*" [69].

9. The semantic structure of the genetic code

The most consequent and systemic discrete analysis of genetic code is given in Shcherbak [84–91]. For the first time an attempt is made to analyze the structure of the genetic code not from widely accepted physicochemical backgrounds, e.g., by looking for interactions between nucleotides and amino acids and their biochemical pathways, but by trying to reveal the semantic structure of the genetic code.

Lets focus again on the standard genetic code table (Fig. 1). The table consists of 16 blocks, with four triplets in each. Eight blocks in this table contain four synonymous triplets, that is, each of these blocks can be considered degenerated and represents a single appropriate amino acid: Gly, Ala, Ser, Pro, Val, Thr, Leu, and Arg. Each of the remaining 8 blocks of the table maps either two amino acids, or an amino acid and the stop-triplet. Except for the property indicated, these two groups of blocks are bound by Rumer's transformation [54,78], indicating that under substitution of all purines to pyrimidines in accordance with the rule TCAG → GACT, the first group is transformed into the second and vice versa. This partitioning of the table into two blocks is the only one, which meets Rumer's rule with Shcherbak

[4] Note that in the papers Duplij and Duplij [31,32] numbers are found that to certain limits may refer to magic numbers, i.e., the sum of diagonal matrix elements of doublets equals to 20, being the number of canonical amino acids (see above), although neither amino acids, nor their number were analyzed by the authors.

[85,86,90] claiming that the ratio $R = \sum(C+G)/\sum(A+T) = 3$ is valid both for the first and the second nucleotide positions in the triplets of the first group. For the second group of triplets, this ratio is naturally reversed, that is, $R = 1/3$. These regularities of the genetic code were found first by Rumer [78], but in the succeeding 20 years they were considered mostly as accidental coincidence. After detailed analysis made by Shcherbak in 1988–1989, the situation has drastically changed. Shcherbak [90,91] revealed fundamental relationships between genetic code structure and nucleon structure of nucleuses of chemical elements composing 20 canonical amino acids.[5]

Since each chemical element is represented in nature as a set of stable isotopes, its molecular mass is always non-integral. However, if we suppose, following the argument by Shcherbak [85,86], that amino acids constituting the proteins are built only from generally occurring natural isotopes, i.e., ${}_1H^1$ (99.9852), ${}_6C^{12}$ (98.892), ${}_7N^{14}$ (99.635), ${}_8O^{16}$ (99.759), ${}_{15}P^{31}$ (100), and ${}_{16}S^{32}$ (95.0) (in brackets, content of an isotope occurring in nature is given in atomic percents), then atomic mass will be expressed as an integer. This supposition is based on plants that were cultivated in the presence of ${}_6C^{12}O_2$ and ${}_6C^{14}O_2$, and selectively assimilated carbon dioxide containing the light isotope ${}_6C^{12}$ [15]. The physicochemical mechanism of this selection is not clear, because from the chemical viewpoint both isotopes are equivalent. Currently, this empirical rule, or *assortative selection by living organisms of only and only one isotope, ${}_6C^{12}$, out of all naturally occurring isotopes, ${}_6C^{12}$, ${}_6C^{13}$, and ${}_6C^{14}$*, is used by biochemists for detection of the share of "biological carbon" in soil samples [1,15].

As known, canonical amino acids are composed of standard peptide groups, with atomic mass, expressed as an integer, equaling to 74, and by side chains with atomic masses varying within the range from 1 for Gly to 130 for Trp. The summarized atomic mass of the side chains of amino acids entering the second 8-blocks group equals 1110. This group consists of 15 amino acids, with summarized mass of their standard peptide groups equaling to $74 \times 15 = 1110$ (Fig. 2). This exact coincidence of two large numbers could hardly be considered accidental, because other regularities exist in each of the eight selected blocks of the genetic code table [91].

For amino acids encoded by a group of triplets, entering the first eight blocks, such ideal coincidence of numbers is not observed. However, some arithmetical curiosities can be found here, too. The sums of atomic masses of peptide groups and side chains for the whole group of amino acids equal to 333 and 592, respectively. Then, the whole sum of amino acid atomic masses in this group can be calculated by adding 333 to 592, which equals 925. The least common multiple for the numbers entering this arithmetical equation equals to $PQ = 37$ (by Shcherbak *Prime Quantum*). Dividing the equation by this number, we arrive at $9 + 16 = 25$ or $3^2 + 4^2 = 5^2$ (see Fig. 2). Is this Pythagorean correlation accidental or does it have a deeper meaning? One might speculate that this pattern may be related to living systems existing in Euclidean space.

Numerical analysis of nucleotide sequences composing DNA [51] and amino acid sequences composing proteins [30] demonstrated that a sort of "mass balance" characterizes these polymers. In particular, molecular masses (or nucleon content, by Shcherbak) of AT and GC pairs of DNA equal to 259 and 260, respectively, so that for rather extended regions, the density or the number of nucleons per a single nucleotide stays almost constant.

[5] Note that the similar attempt was made in the Negadi papers of 2002–2003.

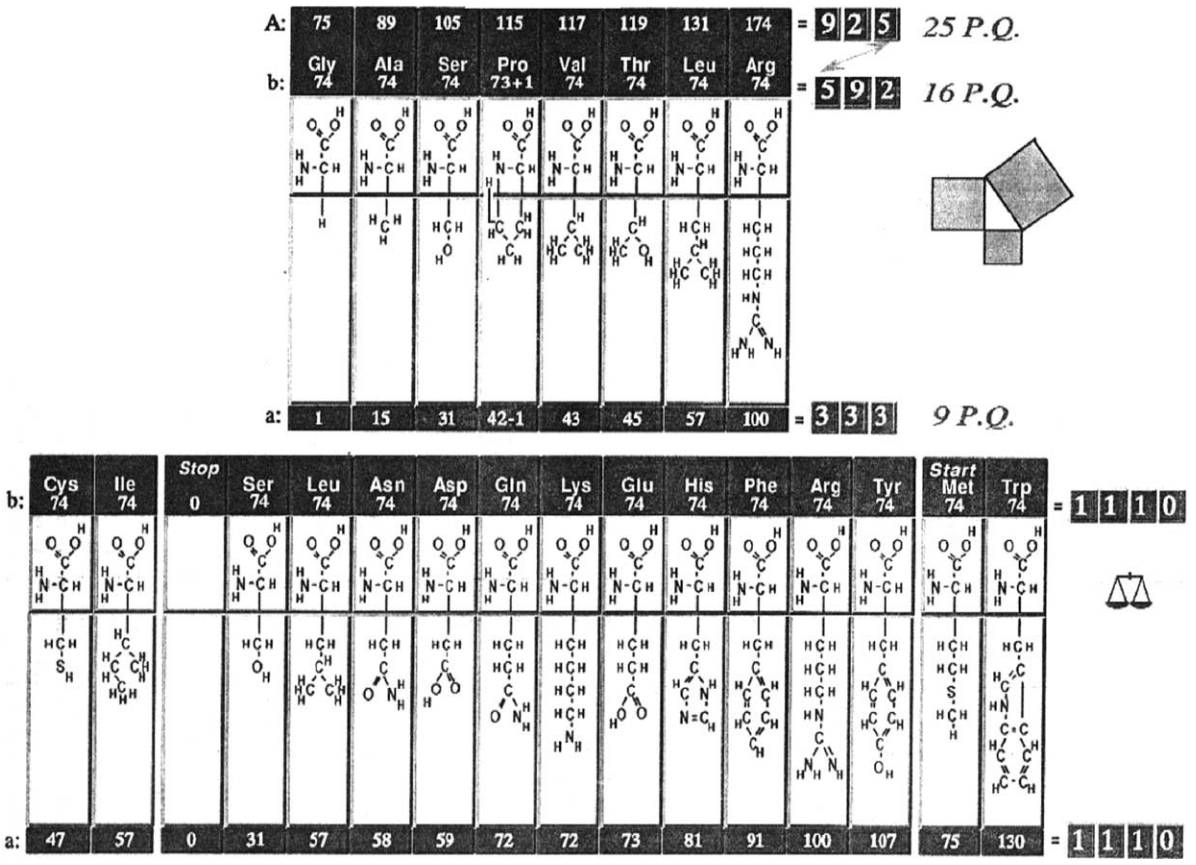

Fig. 2. Algebraic group correlations between atomic masses of amino acids. Here $PQ = 37$. (PQ = Prime Quantum, printed by courtesy of Shcherbak).

In the study made by Downes and Richardson [30], the protein structure of more than 200 pro- and eukaryotic genomes was analyzed. The average density of side amino acid chains was equal to the mass of a standard peptide group with variation not exceeding 10–15%.

The rigorous numerical balance of masses inherent to the genetic code structure is disrupted in real genomes. However, this disruption stays "within the limits of error". Thus, we cannot exclude that it carries some additional information.

It is difficult to imagine that the regularities found by Shcherbak[6] in genetic code structure are not related to physicochemical properties of amino acids and nucleotides. Shcherbak has obtained numerical correlations by analyzing nucleon content of nucleuses and chemical elements making up amino acids. However, he did not relate his findings to physicochemical structure, which is determined by electron surroundings. Also, neither polarity of amino acids, nor the volume, nor hydrophilicity/hydrophobicity properties were taken into account in his analysis. Thus, the question arises: what are the reasons of

[6] The web-site is Internet-available, where the author has detected a series of analogous numerical regularities and symmetries [14].

the numerological properties of the genetic code? One may note that the heuristic approach applied by Shcherbak is somehow similar to that of D.I. Mendeleev. Indeed, periodical properties of chemical elements were discovered by Mendeleev after rearranging the elements in order of increasing atomic weights (masses) and dividing them into groups. Shcherbak applied the same criterion of ordering and determined numerological properties of the groups (Fig. 2). As given in Shcherbak ([91], Figs. 24, 25) and Shcherbak ([91], Figs. 12–14), which are symbolical illustrations of amino acids as the triplets of bases, the respective calligrammes (the nucleotide sequences) are characterized either by bilateral symmetry relative to the central axis of the groups, or by complimentary symmetry. In the first case, the nucleotide sequences completely coincide while reading them to the right and to the left from the central axis of a group. In the second case, such coincidence is achieved with accuracy to complimentarity of the nucleotides. For example, the sequence of bases in the second triplet position in the quasi-group III-II-I (see Fig. 2) has the following sequence: GTAGTAAAAAATGATG, that is, bilateral symmetry is clearly seen. In the quasi-group IV, the sequence of bases in the first triplet position is GGTCGACC. Here we see the symmetry relatively to the central axis of a sequence with accuracy up to being complimentary. We have shown here only a small number of examples extracted from the pool of analogous symmetries represented in Shcherbak [84–91].[7]

These properties of codon sub-groups and the genetic code as a whole may cause an amazement and aesthetic understanding of the genetic code on one hand, but on the other hand may provide evidence that the method of analysis chosen is really constructive. As already Albert Einstein noted: *The ancient science, arithmetic, became one of the powerful tools for guessing the greatest secret of Nature*.

As was noted above, any symmetry is observed only if it reflects some law of conservation [33]. Any algebraic or differential equation relating several variables means that the local law of conservation for these variables is met. By using group properties of symmetry for various codon combinations, Shcherbak developed the complete system of linear algebraic Diophantine equations. The solutions of this system are the integers corresponding to molecular masses of 20 canonical amino acids, whereas zero value, naturally, corresponds to the stop-codon.

Thus, Shcherbak's findings are not that surprising, because in all equations that relate combinations of various amino acids, the numerical values of their masses (or nucleon content, are included a priori. A posteriori, it is always possible to compile a variety of combinations out of 23 "unknown quantities" (for 20 amino acids and three stop-codons) into a set of linear equations, with solutions as respective numerical values. However, it should be noted that the ordering of the self-consistent system of Diophantine equations was not provided accidentally, but it was grounded on the principle of codon group symmetry. From this viewpoint, the integer solutions of this system, corresponding to the canonical set of amino acids, serve as a sort of inner control of the group method. These solutions provide some evidence for non-trivial patterns in the genetic code's group symmetry: "*Arithmetical regularity is revealed by the standard procedure that sets canonical amino acids and the stop codon into functional dependency upon particular parameters of their triplets. In accordance with the rules of procedure, the triplets, first, by means of their properties, order the logics of dividing the code into groups, whereas amino acids, in a subordinate way, demonstrate inside these groups the known properties of balanced nucleon sums and special recording in decimal numeration. Starting from supposition about general inner symmetry of the code, we may expect that there exists a supplemental regularity that compensates for the resulting asym-*

[7] Note that under matrix representation of the genetic code, the authors [31,32] and Negadi [68–71] have also observed various types of symmetry, both in symbolic and numerical representations.

metry: *the logic of ordering is now dictated by the parameters of amino acids and triplets should follow it*". *Further on*: "*It should be emphasized that disintegration into elementary fragments is preconditioned not only by the physicochemical properties of triplets, but, primarily, by the laws of algebra, which are not less solid under formal ordering of genetic code. They inevitably come into force as it becomes clear that arithmetical regularity and cooperative symmetry are real properties of the code*" [91].

Relying on this statement, we may establish that algebraic symmetries that were observed and studied by authors [31,32,48,49,68–71], but were not carefully interpreted, have their background in the arithmetical symmetry of the code. This algebraic layer of symmetries of the genetic code is still waiting for its explorers.

10. Life as a biological computer

One cannot escape the existing similarity to the elementary particles and elements in the universe. These elements and properties were determined by the creation of the observed universe.[8] The chemical structure of the four nucleotides, the 20 canonical amino acids, and the transformation of the set of triplets onto the set of amino acids may be pre-determined as well. Or, the similarity may be the result of an interplay between a few fundamental physical laws. Either way, life would be a fundamental characteristic of the universe. The properties of the genetic code listed above might be sufficient[9] for hypothesizing that life is intrinsic to our universe and belongs to the list of fundamental laws of Nature. If so, the genetic code acquires the status of a non-interpreted category, like the fundamental physical laws of conservation of energy, momentum, and impetus.

We may suppose that the genetic code is formed not only on the basis of physicochemical interactions between triplets and amino acids in the space or on the basis of biochemical metabolic pathways, but first of all, on linguistic basis (Shcherbak, personal communication), which has a character expression in decimal notation. Thus, we may suppose that all living objects, from microorganisms to human beings, can be compared to some extent to biocomputers at a macromolecular level having different levels of complexity. The idea to consider the microbial cell as a *molecular computer* is discussed in literature beginning from the 1970s [58–61,65]. Below, we cite a part from the paper by Liberman published online at (http://www.iitp.ru/personal/Efim/Liberman/intr.html):

"*The foundation of a science associating physics, mathematics, and biology is based on four principles*: (i) *the least cost of action paid for calculation and measurement*, (ii) *optimal predictability*, (iii) *minimal irreversibility, and* (iv) *causation principle given in a new formulation.*

What is life? Could it be described by biophysics and what is biological information? In this article we would like to try to give unanticipated answers to these questions. The assertion reduces to the following: the world is created to be quantum and undulatory in order living beings could act at a minimum onto the future by making measurements and calculations. However, from this novel viewpoint, without living beings that are able to measure and, based on calculation, to predict the future state of (*the*) *surrounding world, the laws of physics do not exist at all.*

[8] Due to external considerations, the analogous conclusion was made by M. Eigen [12].

[9] To do justice, it should be noted that this statement needs strict formal proof. However, at the current stage of knowledge, its proof cannot be obtained in accordance with the rules of classic logic. So, it remains to appeal to intuition.

The necessity in such approach may be understood only if we take into account not only the influence of measurement onto the state of quantum system [...], *but also the influence of calculation by means of limiting calculation systems* [...]. *The limiting calculation systems should have the elements of the minimal size and spend the minimums of free energy and time for maintaining elementary operation. Since energy and time are not quantized, it was supposed* [...] *that we may minimize the action spent* (*the product of energy to the time*). *This value is named as the cost of action. The calculating elements have the first limit of minimal size, molecular dimensions. It was supposed that the managing system of a living cell is a molecular computer, whereas molecular texts, DNA and RNA, are rearranged by using molecular addresses* [...]".

Therefore, the idea that genetic texts may be "treated" by molecular computers by analogy to electron computers is rather aged, and the fact that many researchers have addressed it documents its constructive origin.[10] In this context, the problem of the origin of the genetic code should be considered not only from positions of physicochemical structural and metabolic correlation between triplets and amino acids, but also from positions of their informational correspondence.

The fundamental meaning of "semantic" interaction and cooperative behavior of molecules was noticed by classical nonlinear thermodynamics [75]: "*In this case, it is advisably to discuss the novel coherence, or the mechanism of* "*communication*" *between molecules. However, the interaction of such a type may appear only* (*under*) *very non-equilibrium conditions. Interestingly, the interaction of this type is widely spread in the living world. Its existence could be taken for the very basis of determination of a biological system*". Further on: "*We begin to understand how, presuming from chemistry, to construct complex structures, complicate forms, including precursors of living matter. In strongly non-equilibrium conditions, very important and unexpected property of the matter is estimated*: *henceforth, the physics has every chance to describe the structures as the types of adaptation of a system to external conditions*".

In other words, even at the level of chemical reactions that are incommensurably simple in comparison to a complex network of metabolic reactions in living systems [21,67], but taking place in very non-equilibrium conditions, a certain attribute of living matter appears, that is, *an adaptation*. Obviously, the latter is possible only under condition that interacting molecular aggregates of a system may correlate their state with conditions of the external environment.

Let us consider the limiting case, or the self-consistent, evolutionary selected network of biochemical reactions, and place a question: "*Could the biochemical metabolism be different*?" Buvet [17] emphasizes eight general types of elementary biochemical processes:

1. Redox reaction:

$$(n = 1 \text{ or } 2)$$

$$\mathrm{Ox}_1 + ne^- + p_1\mathrm{H}^+ \rightarrow \mathrm{Re}\,d_1 + q_1\mathrm{H}_2\mathrm{O},$$

$$\mathrm{Re}\,d_2 + q_2\mathrm{H}_2\mathrm{O} \rightarrow \mathrm{Ox}_2 + ne^- + p_2\mathrm{H}^+,$$

$$\mathrm{Ox}_1 + \mathrm{Re}\,d_2 + (p_1 - p_2)\mathrm{H}^+ \rightarrow \mathrm{Re}\,d_1 + \mathrm{Ox}_2 + (q_1 - q_2)\mathrm{H}_2\mathrm{O}.$$

[10] Currently, the ideas are being developed, how to create molecular computer on the basis of known enzyme systems [18,108].

2. Condensation–hydrolysis:

$$\left.\begin{matrix}\equiv C- \\ \equiv P-\end{matrix}\right\} OH + H \left\{\begin{matrix}-O- \\ -S- \\ -N\prec\end{matrix}\right. \rightarrow \left.\begin{matrix}\equiv C- \\ \equiv P-\end{matrix}\right\} \left\{\begin{matrix}-O- \\ -S- \\ -N\prec\end{matrix}\right. + H_2O.$$

3. Addition–eliminations of C–H to C=O:

$$\succ C{=}O + \equiv C\text{–}H \rightarrow HO{\equiv}C\text{–}C{\equiv}$$

4. Addition–eliminations of A–H (A = heteroatom O, S, N) to C=O:

$$\succ C{=}O + A\text{–}H \rightarrow \succ C \prec^{A}_{OH}.$$

5. Addition–eliminations of A–H to C=C:

$$\succ C{=}C \prec + A\text{–}H \rightarrow \succ CA\text{–}CH\prec.$$

6. Enol-oxo isomerisations:

$$_{O}{\geq}C^{H} \succ C\prec \rightarrow \ _{OH}{\succ}C{=}C\prec.$$

7. Methyl transfers (X, Y = C or A):

$$X\text{–}CH_3 + H\text{–}Y \rightarrow X\text{–}H + H_3C\text{–}Y.$$

8. Isoprenyl pyrophosphate polycondensations.

As a consequence of metabolic pathway analysis, Buvet [17] draws the conclusion: "*This means that the biochemical metabolism is not determined by any random choice of enzymic sequences, or by any choice of enzymic sequences which could have evolved from any closed game between polypeptide and polynucleotide polymers. In fact, on the contrary we should consider that the metabolism is as it is because it could not be different owing to the physicochemical properties of carbonaceous compounds in aqueous media. Evolution has systematically explored all possibilities of this particular chemistry, selecting all possible tools for fulfilling the job as well as possible, starting from the state of the art which have been just reached the day before, and making its choices for the maximum benefit of the whole system, which does not necessarily for the best benefit of each of the involved processes taken separately*".

In other words, the author effectuates a conclusion about deterministic origin, which is present both in a metabolism as it is and in the spectrum of enzymes realizing it. Analysis of metabolic pathways of synthesis of inner-cellular molecular structures out of the simplest organic and inorganic substrates reveals that the work of inner-cellular enzymes is subordinate to a rigid biochemical logic [34,67,72], that is, there exists a subsequent chain of biochemical reactions,[11] that passes in a "well-directed" manner the initial substrates and the products of their transformation inside the pool of the executive enzyme apparatus. As the final result of this process, almost an exact copy of molecular structures called a microorganism appears.

[11] The shunting of spme biochemical pathways do not contradict, but emphasize the strictness of biochemical logic, because they increase stability of the biochemical machinery of the cell.

The chemical basis of Earth living material is formed by six elements, C, O, H, N, P and S, which are the basic constituents from which the whole variety of organic molecules is synthesized. Due to the modern enzyme classification, there exist only six classes of activities: oxidoreductases, transferases, hydrolases, liases, isomerases, and ligases, which support the entire spectrum of biochemical reactions in living systems [67]. Each class of activity contains from three to eight types of chemical reactions, transmitting or modifying various groups of atoms composed by C, O, H, N, P, and S, the complete set of enzyme activities participating in the self-reproduction of a bacterial cell hardly exceeds 30. Therefore, a relatively small variety of catalytic activities is sufficient for maintaining the entire set of chemical reactions supporting the biosynthesis of a bacterial cell (although very different requirements exist for heterotrophic organisms, for examples, compared to photosynthetic or chemoautotrophic organisms).

From this viewpoint, the known allegoric metaphor "DNA knows *what* to do but it is not capable, while proteins are capable, but they do not know *to what*" may be transformed as follows: "biochemical logic determined by the genetic code, as a fundamental law, knows *what* to do, while proteins and DNA know *how*". That is to say, it is postulated that biochemical logic of enzyme reactions is a controlling program of self-reproduction in a cell, while proteins are executing elements of this program, which appeal to DNA in case executing elements necessary at a subsequent stage of synthesis are absent.

11. Homochirality of the structural elements of the genetic code

Let us analyze one more universal property of living material, directly referring to the problem under discussion, i.e., homochirality of macromolecules and their monomers participating in storage, transcription, and translation of genetic information. All structural elements of the living system responsible for its self-reproduction are represented by homochiral isomers: nucleic acids containing only D-sugars and proteins containing only L-amino acids. In bacterial cells, these molecules make up about three-quarter of all organic material. In inorganic materials, on the other hand, mirror isomers of molecules are always represented by a racemate. *Thus, living systems formally ignore the principle of equality of right-handed and left-handed forms, hence, in this sense*; *they breach the law of conservation of the spatial evenness*. In any case, until now, neither cells were found on Earth that contain inverted isomers of L-sugars and D-amino acids into nucleic acids and proteins, respectively. Kizel [52] argues that the chiral preference of life cannot be explained only by accidental choice for constructing L-amino acids and D-sugars at the early stages of evolution of live systems. His statement is based on the observation that in non-living world, the "left-handed and right-handed" molecular forms may co-exist inside one and the same areal extent (for instance, quartz crystals are observed in the right and left modifications within the limits of a single mineral deposit). If living systems appear as an obligate event in course of chemical and prebiological evolution, we have no grounds to give the preference to this or that enanthiomorph a priori. A hypothesis on exclusion by existing living systems' forms of their mirror analogues in the process of biological evolution and imminent competition for the chemical sources of energy is doubtful. At the early stages of biological evolution, such competition simply might have not existed due to immeasurableness between masses of organic and inorganic matter. Instead those organisms might have lived next to each other in friendly co-existence. Assuming that the first organisms were heterotrophs (e.g., [81]), the organisms using D-sugars would have no competition from the organisms using L-sugars.

Thus, on the basis of currently known models of origin and evolution of living systems, it is impossible to give preference to any of two mirror forms of molecules. Whether the preferred enanthiomorphic

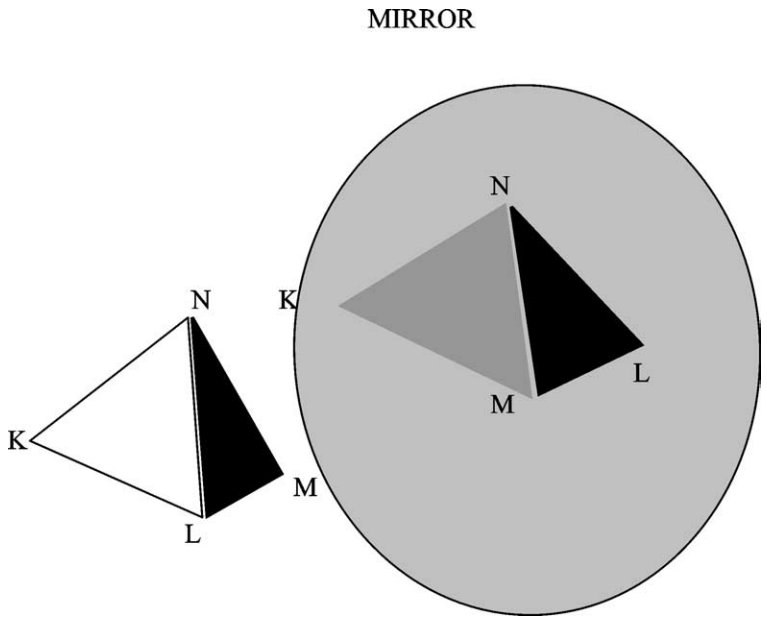

Fig. 3. Tetrahedron (left) and its mirror image (right).

configuration on Earth was only a chance-event, a frozen accident so to speak, or has a deeper physical reasoning, is presently unknown.

From the viewpoint of molecular biology, monochirality of informational macromolecules is required due to geometrical constraints. The atom of carbon, which is the basis for constructing all polymers of living systems on Earth, is ideally suited having four identical covalent bonds directed to the angles of a tetrahedron, so that mirror enanthiomorphs could be formed, each acting with different substituents as denoted in Fig. 3 with the letters KLMN.

Polymers formed out of three such blocks are thermodynamically more stable if the blocks belong to one type of symmetry (all the same, D- or L-conformation), rather than if they would represent a racemic mixture [11,34]. Thermostability is very important for DNA and RNA molecules, the carriers of information.

The process of translation of information, as it is viewed by contemporary biologists, would be much more complicated if the blocks, from which translated and to be translated molecules (i.e., nucleic acids and proteins, respectively) are constructed, were represented by racemic mixture of their monomers.[12] Thus, the phenomenon of monochirality of informational molecules naturally follows from their composition and structure. The particular pattern of chirality of a molecule does not really play the role. Hence, within the frames of molecular-biological viewing, no argument can be formulated for choosing D- or L-enanthiomorphs. Interestingly, experimental physics encountered a somewhat similar problem, the problem of spatial anisotropy of electrons in the process of β-splitting of a nucleus [57,106]. However, theoreticians rather quickly managed to "reseal a defensive line" by appealing to one fundamental

[12] Mechanical illustration of this phenomenon: the right-hand nut may be fastened to a right-hand (but not left-hand!) screw. In the model considered the DNA thread (double spiral) stands for a screw, while transcribing protein complex serves as a nut.

theorem, the so-called **CPT**-theorem by Luders–Pauly [41]. One of the formulations of this theorem states the following [66]: "*for any interaction, the product of three inversions, including operation of charge conjugation* (C), *perversion* (P), *and time reversal* (T), *is invariant*". The CPT-theorem is valid for each theory that complies with the principle of causation. The meaning of discrete rearrangements considered here is as follows:

- under spatial perversion, that is, $\mathbf{P} \to -\mathbf{P}$ condition of particles, with dimensional coordinates **r**, impulse **p**, and spin projection σ (for spins differing from zero), is transformed into the same condition with coordinates equaling to $-\mathbf{r}$, impulse, to $-\mathbf{p}$, while spin projections σ stay invariant;
- under time reflection $\mathbf{T} \to -\mathbf{T}$, condition of particles with coordinates **r**, impulses **p**, and spin projections σ is transformed into the same condition with invariant coordinates **r**, but with impulses equaling to $-\mathbf{p}$, while spin projections σ have an opposite sign; also, time direction is changed to the opposite one, i.e., projectile particles are replaced by escaping particles, and vice versa;
- under charge conjugation **C**, the particles are transformed into corresponding anti-particles without changing of **r**, **p**, and σ.

As follows from the logic of the **CPT**-theorem, the inversion of the physical system or a process occurring in the system, relative to one of these transformations, is associated with inversion of one of the remaining transformations, so that their product stays still invariant. Thus, for the non-conservative **P**-evenness in the processes related to β-splitting, either combination **PT** or **PC** should be invariant. Without entering into details of experimental and theoretical collisions in the sphere of elementary particles, we will try to apply this methodology for analysis of spatial anisotropy of living systems.

For living objects existing on Earth, the multiplier **C** should be considered invariant, because living matter consists of particles, not anti-particles. If so, the product **TP** is obligatory invariant. Hence, it follows that for the living objects, which are the mirror images of those existing on Earth, with inverted space (i.e., $\mathbf{P} \to -\mathbf{P}$), the time should also have an opposite sign, i.e., $\mathbf{T} \to -\mathbf{T}$, then $(-\mathbf{P})(-\mathbf{T}) = \mathbf{PT}$.

The steady statement of the living system is possible only under strict adherence to the casual correlation between biochemical processes of this system. Realization of genetic information in a form of functional protein units may take place only in a strict time ordering: a protein is not synthesized prior to its corresponding RNA. The rigid causative determination of processes makes a demand for the time to be unidirectional. If it is assumed that the CPT—theorem can be "upscaled" from elementary particles to organic macromolecules, then, in accordance to the **CPT**—theorem, it requires chirality of molecules participating in storage and transformation of information. Otherwise, by assuming independent flow of processes occurring in a living system, occurring forward or backward in time, we would accept an existence of racemic mixture of informational molecules in a cell.[13] Thus, we arrive at the conclusion that in living systems, the space, or speaking more exact, the spatial distribution of atoms, and time, that is, the direction of development of molecular processes, are not independent of time.

The condition of unidirectional flow of time, or causative conformity of processes in a living system impose requirements to its structural elements: *informational macromolecules should consist of atoms enabling formation of chiral polymers*. Carbon as the chemical basis of life on Earth thus ideally fits into such a scheme. The monochirality concept in context with the molecular structure of the carriers of

[13] "*The causative relations are realized through the evidence that time is definitely directed, from past to future*" [77].

information provides additional argumentation in favor of the hypothesis stating that the microbial cell is in principle a molecular computer.

12. Discussion

Life is a very particular phenomenon that sharply diverges from all other phenomena in the physically observable domain of experience [73]. This is why a posteriori, we try to formulate what is the subject of study and try to introduce the formalism of a meaningful description of life. At first view, this way of touching upon this question seems absurd due to the "obviousness" of the difference between the life and non-life. However, by trying to define what the exact differences are, this obviousness is transformed into an almost insoluble problem how to formulate the strict formalization of a definition of life.[14]

The problem of the origin of life is more challenging if we impose a restriction on the age of the universe. In case of a universe with unlimited time, any profound creation, including the first microbial cell as the starting point of life, although exceedingly improbably, might in principle appear at once due to an accidental combination of atoms best suited for the creation of these objects. This would be exceedingly unlikely to happen in the life time of our about 14 billion year old universe, but who would say that there were not a zillion universes before our universe? In most or all of the other universes the conditions were not right or simply the chance event did not happen, and no observer was created. This line of argument is often made with the cosmological constants of being just right for the development of matter, evolution of galaxies and planetary systems. An "accidental assembling" of any live object from molecules does not contradict the 2nd Law of Thermodynamics either, it just requires an exceedingly improbable amount of time. Under such a line of thought, the molecular machinery of a living cell can be considered simply as an accidental lucky combination of atoms with the capability of self-replication.

Yet, based on life on Earth there may not be much time required, which would counter the argument just advanced. The earliest evidence of life on Earth is 3.8 billion years ago and these forms are presumably directly ancestral to all groups still alive today with all exhibiting the same genetic code [19]. Thus, the genetic code must have evolved to a stable configuration before that time. With the Earth being about 4.6 billion years old and being uninhabitable in its earliest history (due to formation under hot conditions, the cataclysmic impact that formed the Moon, heavy meteorite bombardments etc.), only a few hundred million years remain at best for the development of the genetic code and life itself. James and Ellington [47] argue that evolution proceeds in a step-wise fashion, and once a stage is achieved, other things then become so much more likely. However, there is also a sense that given a world of DNA and amino acids, then perhaps the genetic code we know is more or less an inevitable outcome [19].

The notion of living matter cannot be understood without determining its outer environment, since these two essences are interdependent. The universe without a (higher) form of life is unobservable: if a sensitive structure does not exist, nobody will put questions and answers forward. The capability of self-reproduction in this context is an essential attribute of an observable universe, because without self-reproduction there is no evolution, for Earth-type life at least. In this connection, physicists put forward the alien-sounding notion of the anthropogenic principle, stating that the fundamental laws of physics,

[14] The absence of a unified definition of life was clearly demonstrated at the conference held in September 2000 in Moden (Italy) [39].

as well as cosmic constants, have to be exactly the same with very little margin of error as they are observed [16,76]. *Only at these specified conditions, life can appear in the universe*.

The task to find answers to what life is and how life originated is the deepest question of our time and considerable resources should be used to find answers to it. However, possibly there is no definition to determine the "real" essence of life. According to Gedel's theorem *neither system is self-conscious within the frames of its own notions* [64,74]. Therefore, there may be no definition comprehensible to us to describe the real essence of live. From a scientific perspective, though, this notion is utterly unsatisfying. And it is not consistent with progress that has been made in the deeper understanding of life. For example, a definition for life, modified from Gusev [42], can be given as:

- a life object as a space-time limited structure, informationally sufficient for reproduction in an adequate environment, that inevitably arises at a particular stage of evolution in the universe;
- a life system as a set of functionally interacting life objects, with different complexity, that are capable to task-oriented modification of the outer environment;
- life as an energetically dependable process of periodical rearrangement of elementary units of a substance, resulting in permanent increase of structural and functional complexity of life systems and their environment.

This definition nicely provides insights into the functionality of life. Other definitions are available that emphasize other aspects such as its detectability (e.g., [82]).

Did we answer the 9 questions put forward in the introduction section of this review? Probably not, but we provided routes chosen by different researchers, that addressed these issues. In our review these issues were addressed by the following authors:

1. Why is genetic code universal [84–91]?
2. Did "genetic dialects" appeared accidentall or by some kind of selection process [84–91]?
3. Why is the genetic code represented by the four bases, A, T(U), G, and C [4,5]?
4. Why does the genetic code have a triplet structure [4,5]?
5. Why is the genetic code not overlapping, that is, why does the translation apparatus of the cell transcribe information as a discrete value equaling three, but not one or two [4,5]?
6. Why does the extent of degeneracy of the code vary from one to six for various amino acids [84–91]?
7. Is the existing distribution of codon degeneracy for particular amino acids accidental or some kind of selection process [31,32,48,49,68–71,84–91]?
8. Why were only 20 canonical amino acids selected for proteins synthesis [31,32,48,49,68–71,84–91]?
9. Is the choice of exactly this set of amino acids accidental or due to some kind of selection process [31,32,48,49,68–71,84–91]?

Paradoxical though it may seem, more recent progress has been made answering these questions by employing physical processes and models rather than biological methods. Probably, the phenomenon of life and its origin cannot be adequately addressed within the frames of biological categories alone. In other words, the fundamental laws of physics that cannot be violated in the observed universe determine not only the structure of the micro- and macro-world, the evolution of inert matter, but they also intrinsically govern living matter. Examples discussed here are the uniformity of the genetic code in Earth biota and monochirality of its informational molecules, proteins and nucleic acids. Since an observer is the

only criterion proving the existence of the universe, we do not have any information about unobserved universes, which cannot be examined experimentally and become abstractions. The origin of the simplest living systems occurred at a particular stage of the universe's evolution, and is thus in accordance with the fundamental laws of physics. From this viewpoint, the "curiosities" of the genetic code described above do not seem accidental or factitious. Just the contrary, they set the researchers to search for in-depth relationships between the laws of the living and non-living nature. The understanding of the origin of life and the universe are interrelated, and it is probably impossible to understand, what the universe is, without understanding beforehand, what life is [62].

13. Epilogue

We discussed the different arguments made for the origin of the genetic code with the goal to provide a balanced discussion. By working and thinking about this material, we developed our own opinions and like to share them with the reader. Based on the reviewed literature, we feel that the arguments made favored a deterministic origin of life compared to an accidental selection of "favorable" combinations of molecules-progenitors of living cells. The majority of authors seem, explicitly or implicitly, tending toward this direction.

Based on the analysis of the phenomenon of life and its attributes, a genetic code or information code, is absolutely essential to any living organism. Life as such and its information code is intrinsically linked to the physical laws of the universe. Thus, without understanding life we cannot understand the universe and vice versa. Whether life is an inevitable outcome of a universe such as ours can be disputed, but it is a well-grounded idea.

Life is based on the molecular machinery with which we are familiar for life forms on Earth. Any extraterrestrial life would require such a complex machinery as well. The element that has an extreme versatility and allows for such a complex machinery is carbon, thus most or all life anywhere else in the universe is likely to be carbon-based as well. The design of existing carriers of information (DNA and RNA) comprised of A, T(U), G, and C bases, and their functional executors (proteins) constructed from a set of 20 canonical amino acids works extremely well for life on Earth and is extremely optimized. Whether this represents only one of the best or the best solution to life is unclear, but we should not be surprised if we find extraterrestrial life in the future that has the same or a very similar molecular machinery.

Acknowledgements

We thank Professor Nikolay Kolchanov for helpful discussions and also an anonymous reviewer for his constructive input. This work was supported by the Foundation of Integration Programs of SB RAS (grant No. 148, 2003).

References

[1] Abelson PH, Hoering TC. Carbon isotope fractionation in formation of amino acids by photosynthetic organism. Proc. Natl. Acad. Sci. 1961;47:623–32.

[2] Alberti S. The origin of genetic cod and protein synthesis. J. Mol. Evol. 1997;45:352–8.
[3] Alberti S. Evolution of the genetic code, protein synthesis and nucleic acid replication. Cell Mol. Life Sci. 1999;56:85–93.
[4] Aldana M, Cazarez-Bush F, Cocho G, Martinez-Mekler G. Primordial synthesis machines and the origin of genetic code. Physica A 1998;257:119–27.
[5] Aldana-Gonzales M, Cocho G, Larralde H, Martinez-Mekler G. Translocation properties of primitive molecular machines and their relevance to the structure of the genetic code. J. Theor. Biol. 2003;220:27–45.
[6] Amirnovin R. An analysis of the metabolic theory of the origin of the genetic code. J. Mol. Evol. 1997;44:473–6.
[7] Ardell DH. On error minimization in a sequential origin of the standard genetic code. J. Mol. Evol. 1998;47:1–13.
[8] Ardell DH, Sella G. On the evolution of redundancy in genetic codes. J. Mol. Evol. 2001;53:269–81.
[9] Ardell DH, Sella G. No accident: genetic codes freeze in error-correcting patterns of the standard genetic code. Phil. Trans. R. Soc. Lond. B. 2002. DOI 10.1098/rstb.2002. 1071.
[10] Arzamastsev AA. The nature of optimality of DNA code. Biophys. Russ. 1997;42:611–4.
[11] Avetisov VA, Goldansky VI. Physical aspects of breaking the mirror symmetry of bioorganic world. Achiev. Phys. Sci. Russ. 1996;166:873–91.
[12] Babloyantz A. Molecules, dynamics, and life. New York: Willey & Sons; 1989.
[13] Balasubramanian R, Seetharamulu P, Raghunathan G. A conformational rational for the origin of the mechanism of nucleic acid-directed protein synthesis of 'living' organisms. Origins Life 1980;10:15–30.
[14] Boulay: http://perso.wanadoo.fr/jean-yves.boulay/rap/eng.htm.
[15] Calvin M. Chemical evolution. Oxford: Clarendon Press; 1969.
[16] Carr BJ, Rees MY. The anthropic principle and the structure of the physical world. Nature 1979;278:605–12.
[17] Buvet R. Could the biochemical metabolism be different? Origin of Life. In: Wolman Y, editor. Dordrecht: Reidel; 1981, p. 589–99.
[18] Conrad M, Zauner K-P. Conformation-driven computing: a comparison of designs based on DNA, RNA, and protein. Supramol. Sci. 1998;5:787–90.
[19] Conway MS. Life's solution: inevitable humans in a lonely universe. Cambridge, UK: Cambridge Univ. Press; 2003.
[20] Crick FHC. The origin of genetic code. J. Mol. Biol. 1968;38:367–79.
[21] Dagley S, Nicholson D. An introduction to metabolic pathways. Oxford: Blackwell; 1970.
[22] Danckwerts HJ, Neubert D. Symmetries of genetic code-doublets. J. Mol. Evol. 1975;5:327–32.
[23] Di Giulio M. The β-sheets of proteins, the biosynthetic relationships between amino acids, and the origin of the genetic code. Origins Life Evol. Bioph. 1996;26:589–609.
[24] Di Giulio M. On the origin of genetic code. J. Theor. Biol. 1997;191:573–81.
[25] Di Giulio M, Medugno M. The historical factor: the biosynthetic relationships between amino acids and their physicochemical properties in the origin of the genetic code. J. Mol. Evol. 1998;46:615–21.
[26] Di Giulio M, Medugno M. Physicochemical optimization in the genetic code origin as the number of codified amino acids increases. J. Mol. Evol. 1999;49:1–10.
[27] Di Giulio M, Medugno M. The level and landscape of optimization in the origin of the genetic code. J. Mol. Evol. 2001;52:372–82.
[28] Di Giulio M. A blind empiricism against the coevolution theory of the origin of the genetic code. J. Mol. Evol. 2001;53:724–32.
[29] Di Giulio M. Genetic code origin: are the pathways of type Glu-tRNA^{Gln} → Gln-tRNA^{Gln}. J. Mol. Evol. 2002;55:616–22.
[30] Downes AM, Richardson BJ. Relationships between genomic base content and distribution of mass in coded proteins. J. Mol. Evol. 2002;55:476–90.
[31] Duplij D, Duplij S. Determinative degree and nucleotide content of DNA strands. Biophys. Bull. 2000;497:1–7.
[32] Duplij D, Duplij S. Analysis of symmetries of genetic code and the extent of determination of codons. Biophys. Bull. Kharkov State Univ. 2000;488:1–11 (in Russian).
[33] Elliott JP, Dawber PG. Symmetry in physics. London: Macmillan; 1979.
[34] Elliott WH, Elliott DC. Biochemistry and molecular biology. Oxford: Oxford Univ. Press; 1997.
[35] Fitch W, Upper K. The phylogeny of RNA sequences provides evidence for ambiguity reduction in the origin of the genetic code. Cold Spring Harbor Symp. Quant. Biol. 1987;52:759–67.
[36] Freeland SJ, Laurence DH. The genetic code is one in a million. J. Mol. Evol. 1998;47:238–48.

[37] Freeland SJ, Hurst LD. Load minimization of the genetic code: history dose not explain the pattern. Proc. R. Soc. London Ser. B Biol. Sci. 1998;265(1410):2111–9.
[38] Freeland SJ, Knight RD, Landweber LF, Hurst LD. Early fixation of an optimal genetic code. Mol. Biol. Evol. 2000;17(4):511–8.
[39] Fundamental of Life. 2001. Palyi G, Zucchi C, et al., editors. Elsevier, Paris;
Gamow G. Possible relation between deoxyribonucleic acid and protein structures. Nature 1954;13:318.
[40] Grantham R. Amino acid difference formula to help explain protein evolution. Science 1974;185:862–4.
[41] Grovert G, Luders G, Rolnik G. CPT theorem and its application. Review. Achiev. Phys. Sci. Russ. 1960;71:289–325.
[42] Gusev VA. Living Universe. In: Palyi G, Zucchi C, Caglioti L, editors. Fundamentals of Life. Amsterdam: Elsevier; 2002, p. 545–52.
[43] 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. Nucleic Acids Res. 1996;24:4249–55.
[44] Ikehara K. A possible evolutionary pathway of the genetic code deduced from SNS hypothesis. Viva Origino 1998;26:311–20.
[45] Ikehara K, Omori Y, Arai R, Hirose A. A novel theory on the genetic code: a GNC-SNS hypothesis. J. Mol. Evol. 2002;54:530–8.
[46] Ikehara K. Origins of gene, genetic code, protein and life: comprehensive view of life systems from a GNC-SNS primitive genetic code hypothesis. J. Biosci. 2002;27:165–86.
[47] James KD, Ellington AD. The search for missing links between self-replicating nucleic acids and the RNA world. Origins Life Evol. Biosph. 1995;25:515–30.
[48] Jimenez-Montano MA, de la Mora-Basanez CR, Poschel T. The hypercube structure of the genetic code explains conservative and non-conservative amino acid substitutions in vivo and in vitro. BioSystems 1996;39:117–25.
[49] Jimenez-Montano MA. Protein evolution drives the evolution of the genetic code and vice versa. BioSystems 1999;54:47–64.
[50] Junck JR. The genetic code as a periodic table. J. Mol. Evol. 1978;11:211–24.
[51] Kasharov VV, Krassovitskiy AM, Mamleev VSh, Shcherbak VI. Random sequences of proteins are exactly balanced like the canonical pairs of DNA. In: 10th ISSOL Meeting, 13th International Conference on the Origin of Life, Oaxaca City, Mexico, June 30–July 4. 2002.
[52] Kizel VA. Physical reasons of dissymmetry of living systems. Moscow: Nauka; 1985 (in Russian).
[53] Knight RD, Freeland SJ, Landweber LF. Selection, history and chemistry: The three faces of the genetic code. Trends Biochem. Sci. 1999;24(6):241–7.
[54] Konopelchenko BG, Rumer YuB. The codon classification inside genetic code. Dokl. Akad. Nauk SSSR 1975;223:471–4 (in Russian).
[55] Lacey JC Jr., Mullins DW. Experimental studies related to the origin of the genetic code and process of protein synthesis—a review. Origin Life 1983;13:3–42.
[56] Lacey JC Jr., Wickramasinghe NSMD, Cook GW. Experimental studies on the origin of the genetic code and the process of protein synthesis: a review update. Origins Life. Evol. Biophys. 1992;22:243–75.
[57] Lee TD. Weak interactions and nonconservation of evenness. Achiev. Phys. Sci. Russ. 1958;66:89–97 (in Russian).
[58] Liberman EA, Minina SV, Shklovsky-Kordy NE. BioSystems 1989;22:135, http://www.iitp.ru/personal/Efim_Liberman/int_r.html.
[59] Liberman EA. Molecular computational machine MCM of a cell. General considerations and hypotheses. Biophysics 1972;17:932–43 (in Russian).
[60] Liberman EA. Studying of diffusing modeling system of molecular computational machine in a neuron. Biophysics 1980;25:455–61 (in Russian).
[61] Liberman, E.A., http://www.iitp.ru/personal/Efim_Liberman/int_r.html.
[62] Linde AD. Elementary-particle physics and inflationary cosmology. Moscow: Nauka; 1990 (in Russian).
[63] Maddox J. The genetic code by numbers. Nature 1994;367:111.
[64] Mathematical encyclopedia, Sovetskaya encyclopedia. Moscow; 1977 (in Russian).
[65] Minina SV, Liberman EA. Access and outlet ports of the quantum computer. Biophysics 1990;35:132–5 (in Russian).
[66] Mukhin KN, Tikhonov VN. Old and modern exotics in the world of elementary particles. Achiev. Phys. Sci. Russ. 2001;171:1201–51 (in Russian).

[67] Musil J, Novakova O, Kunz K. Biochemistry in schematic perspective. Prague: Avicenum Czechoslovak medical press; 1981.
[68] Negadi, T., 2002. Cracking the genetic codes with a modular determinative degree. An Algebraic Approach. PREPRINT LPTO/negadi/NoD3101/14/No01/2002.
[69] Negadi T. From the Schrödinger equation to the genetic code and beyond. In: 24th Meeting of the International Colloquim on Groop Theoretical Methods in Physics, Paris, July 15–20, 2002.
[70] Negadi, T., 2003. On the Symmetries of the 16 genetic code-doublets. PREPRINT LPTO/negadi/NoD3101/14/No02/2003.
[71] Negadi T. Rumer's transformation, in biology, as the negation, in classic logic. Int. J. Quant. Chem. 2003;94:65–82.
[72] Nicholson DE. Metabolic pathways. Bucks, England: Koch-Light Laboratories LTD. Colnbrook; 1968.
[73] Palyi G, Zucchi C, Hajdu C, 2000. Theories on the origins of life. Atti e Memorie Acc. Naz. Sci. Lett. Atti Modena. ser. VIII. vol. II, pp. 389–415.
[74] Philosophic encyclopedia, Sovetskaya encyclopedia. Moscow; 1960 (in Russian).
[75] Prigogine I, Stengers I. Order out of chaos. London: Heinemann; 1984.
[76] Rozental IL. Elementary particles and structure of the universe. Moscow: Nauka; 1984 (in Russian).
[77] Rozental IL. Geometry, dynamics, universe. Moscow: Nauka; 1987 (in Russian).
[78] Rumer YuB. On systematization of codons in genetic code. Dokl. Akad. Nauk SSSR 1966;167:1393–5 (in Russian).
[79] Rumer YuB. Systematization of codons in genetic code. Dokl. Akad. Nauk SSSR 1968;183:225–6 (in Russian).
[80] Schrödinger E. What is Life? Minde and Matter. Cambridge: Cambridge Univ. Press; 1944.
[81] Schulze-Makuch D, Irwin LN. Life in the universe: expectations and constraints. Heidelberg, Germany: Springer-Verlag; 2004.
[82] Schulze-Makuch D, Irwin LN, Guan H. Search parameters for the remote detection of extraterrestrial life. Planet. Space Sci. 2002;50:675–83.
[83] Sella G, Ardell DH. The impact of message mutation on the fitness of a genetic code. J. Mol. Evol. 2002;54:638–51.
[84] Shcherbak VI. The co-operative symmetry of the genetic code. J. Theor. Biol. 1988;132:121–4.
[85] Shcherbak VI. The Rumer's rule and transformation in the context of the co-operative symmetry of the genetic code. J. Theor. Biol. 1989;139:271–6.
[86] Shcherbak VI. Ways of wobble pairing are formalized with the co-operative symmetry of the genetic code. J. Theor. Biol. 1989;139:277–81.
[87] Shcherbak VI. The "START" and "STOP" of the genetic code. Why exactly ATG and TAG, TAA? J. Theor. Biol. 1989;139:283–6.
[88] Shcherbak VI. The symmetrical architecture of the genetic code systematization principle. J. Theor. Biol. 1993;162:395–8.
[89] Shcherbak VI. Twenty canonical amino acids of the genetic code: the arithmetical regularities. Part I. J. Theor. Biol. 1993;162:399–401.
[90] Shcherbak VI. Sixty-four triplets and 20 canonical amino acids of the genetic code: the arithmetical regularities. Part II. J. Theor. Biol. 1994;166:475–7.
[91] Shcherbak, V.I., 1995. Mathematics model of the universal genetic code. PhD Thesis. Almaty (in Russian);
Shcherbak VI. Arithmetic inside the universal genetic code. BioSystems 2003;70:187–209.
[92] Shimizu M. Molecular basis for the genetic code. J. Mol. Evol. 1982;18:297–303.
[93] Smith JM, Szathmary E. The major transitions in evolution. Oxford, New York, Tokyo: Oxford Univ. Press; 1995. p. 81.
[94] Szathmary E. Four letters in the genetic alphabet a frozen evolutionary optimum?. Proc. R. Soc. London Ser. B Biol. Sci. 1991;245(1313):91–100.
[95] Szathmary E. What is the optimum size for the genetic alphabet? Proc. Natl. Acad. Sci. USA 1992;89(7):2614–8.
[96] Taylor FJR, Coates D. The code within the codons. BioSystems 1989;22:177–87.
[97] Watson JD, Crick FHC. A structure for deoxyribose nucleic acid. Nature 1953;171:737–8.
[98] Weber AL, Lacey Jr JC. Genetic code correlations: amino acids and their anticodon nucleotides. J. Mol. Evol. 1978;11:199–210.
[99] Weber AL, Miller SL. Reasons for the occurrence of the twenty coded protein amino acids. J. Mol. Evol. 1981;17:273–84.
[100] Woese CR. On the origin of the genetic code. Proc. Natl. Acad. Sci. USA 1965;54:1546–52.
[101] Woese CR, Dugre DH, Dugre SA, Kondo M, Sexinger WC. On the fundamental nature and evolution of the genetic code. Cold Spring Harbor Symp. Quant. Biol. 1966;31:723–36.

[102] Woese CR. The genetic code. New York: Harper & Row; 1967.
[103] Wong JT. A coevolution theory of the genetic code. Proc. Natl. Acad. Sci. USA 1975;72:1909–12.
[104] Wong JT. Role of minimization of chemical distances between amino acids in the evolution of the genetic code. Proc. Natl. Acad. Sci. USA 1980;77:1083–6.
[105] Wong JT. Evolution of the genetic code. Microb. Sci. 1988;5:174–82.
[106] Yang CN. The law of conservation of evenness and other symmetries. Achievements Phys. Sci. Russ. 1958;66:79–87 (in Russian).
[107] Yarus M. Amino acids as RNA ligands: a direct-RNA-template theory for the code's origin. J. Mol. Evol. 1998;47:109–17.
[108] Zauner K-P, Conrad M. Molecular approach to informal computing. Soft Comput. 2001;5(1):39–44.