

Genetic code: Chemical Distinctions of Protein Amino Acids

Miloje M. Rakočević

Department of Chemistry, Faculty of Science, University of Niš, Serbia
(E-mail: mirkovmile@mts.rs; mirkovmiloje@gmail.com; www.rakocevcode.rs)

Abstract. In the work it is shown that 20 protein amino acids ("the canonical amino acids" within the genetic code) appear to be a whole and very symmetrical system, in many ways, all based on strict chemical distinctions from the aspect of their similarity, complexity, stereochemical and diversity types. By this, all distinctions are accompanied by specific arithmetical and algebraic regularities, including the existence of amino acid ordinal numbers from 1 to 20. The classification of amino acids into two decades (1-10 and 11-20) appears to be in a strict correspondence with the atom number balances. From the presented "ideal" and "intelligent" structures and arrangements follow the conclusions that the genetic code was complete even in prebiotic conditions (as a set of 20 canonical amino acids and the set of 2+2 pyrimidine / purine canonical bases, respectively); and the "evolution" of the genetic code can only mean the degree of freedom of standard genetic code, i.e. the possible exceptions and deviations from the standard genetic code.

Keywords. Genetic code; Protein amino acids; Canonical amino acids; Atom number balance; Ordinal amino acid number.

From a chemical point of view the first step of classification of protein amino acids (AAs), "the canonical amino acids" within genetic code, must be the classification into aliphatic and aromatic AAs, where on a hierarchical scale of changes by similarity and complexity, aliphatic AAs must precede the aromatic. For the same reason of the chemical hierarchy, within the class of aliphatic AAs at the beginning must be the hydrocarbon AAs (possessing in the side chain carbon and hydrogen, or hydrogen only, in the case of glycine), and at the end two sulfuric AAs, quite different from preceded non-sulphuric AAs. This means that two sulfur AAs [as last in class of aliphatic AAs] must be found in a direct contact to the aromatic ones.

In further course of sequencing of AAs, in terms of changes by similarity, from the aspect of the AAs singlets and/or doublets, i.e. pairs, it should be considered appropriate distinctions in three areas: in the hydrocarbon, aromatic, and that between them. In the set of aromatic AAs, Phe came first, as the simplest, followed by Tyr, and Trp, all three with possession of a benzene ring. At the very end ultimately must be His, the only one who does not possess the aromatic benzene ring. In the set of hydrocarbon AAs, at the very beginning must be Gly as the simplest one. Follows Ala as the first possible case of hydrocarbon series with an open carbon chain. Then come Val and Pro, both with three carbon atoms in the side chain, rather than Leu and Ile with four carbon atoms. By this, Val with half-cyclic chain precedes Pro with cyclic one; also Ile precedes Leu, as more similar with Pro. [The details of the relationship between Val and Pro, see in: Rakočević & Jokić, 1996; on the relationship between Pro and Ile, in: Rakočević 1998, Survey 4, p. 290, where the pair Pro-Ile is one of ten pairs of amino acids classified into two classes: class I (with Ile), handled by class I of enzymes aminoacyl-tRNA synthetases, and class II (with Pro), handled by class II of enzymes aminoacyl-tRNA synthetases.]

Finally, it remains to determine the chemical distinctions of AAs in "between" area. We have already said that sulfur amino acid pair, Cys-Met, precedes aromatic amino acids. As chalcogene

AAs, they must be in contact with other two chalcogene amino acids, Ser-Thr. By this, the contact have to be made via Cys because it possesses SH group, correspondent to OH group in Ser as well as in Thr.

It is to be understood that a pair of oxygen AAs with the hydroxyl (OH) functional group in side chain must be in contact with a pair of two also oxygen AAs, but which possess the carboxyl (COOH) functional group: Asp-Glu. But the problem is that both of these AAs have their amide derivatives (Asn-Gln) and it is not easy, when determining the distinctions, determine which here precede and which ones follow.

It turns out, however, that the problem easier to solve when returning to the beginning, in the area of hydrocarbon AAs, to the "point" of the pair Ile-Leu. Further must follow the pair of nitrogen derivatives, Lys-Arg, and Lysine must come first with four carbon atoms in the side chain, which number is also valid for Leucine; and then, with the validity of both principles – the continuity and minimum of change – comes Arginine with three atoms (not counting carbon atom in the guanidino group). Then, chemically speaking, it is very natural that after Arginine comes Gln with its precursor, the glutamic amino acid, both (Gln-Glu) with two carbon atoms in the side chain; it is naturally indeed that, in terms of chemical similarity, after 3C atoms occurs changes into 2C atoms, better than into 1C atom, as the case we have in the pair Asn-Asp. [As in the case of the guanidino functional group in arginine, no carbon atom is counting in the carboxylic or amide functional group.] With this, chemical sequencing of series of 20 AAs closes, starting from the first, glycine, and ending with very different histidine.

*

The main result of this pure chemical sequencing of AAs, presented in Table 1, shows that these chemical distinctions are accompanied by specific arithmetical regularities, including the existence of amino acid ordinal numbers from 1 to 20, with two decades (1-10 and 11-20) ; and also shows the full balance of the number of atoms in the 20 amino acid molecules: 102 ± 0 atoms in two decades, as well as on two zig-zag lines, where such a system with two zig-zag lines represents the first possible periodic system with two periods.

However, the result of the most surprising, is the result shown in Table 2. If we take the four by four AAs from Table 1 and line up they in five rows, we obtain an arrangement of AAs, as it is shown in Table 2 where the difference ratio in the number of atoms, per rows equals 11:11:11 [(61-50 = 11); (42-31 = 11), (31-20 = 11)]. But the surprise comes only with an insight into Table 3 in which, exactly in the 11th step of a specific number arrangement where we find the same arithmetical result (20-31-42-61-50) valid for the number of atoms in the arrangement of AAs in Table 2. [Notice that this sequence is in contact with two sequences which contain the first pair of friendly numbers, 220-284, what is a coincidence only, or what more?]¹ However, knowing a previous result (Rakočević, 2011a,b) where the number of atoms in four diversity types of the protein AAs is identical to the unique arithmetical result in a unique arithmetical

¹ A hypothesis on the determination of the genetic code with the perfect and friendly numbers we have presented in the book (Rakočević, 1997b). (www.rakocevcod.rs)

arrangement (Remark 1), surprises for the presented atom number sequence no longer; simply, everything is brought into a connection with the three endpoints of this arrangement: 00-11-22, in the form of a mirror mapping (00-11-22 / 22-11-00) with compression in the mirror plane (00-11-22-11-00).

Remark 1. There is a unique arrangement of the natural numbers series, such as in the "Table of minimal adding" (TMA), with changes in the values for 01 horizontally and just for 11 vertically [01, 02, ..., 09, 10, 11 / 12, 13, ..., 20, 21, 22 / 23, ..., 99] (Rakočević, 2011b, Table 4, p. 826). Moreover, in addition to horizontal and vertical arrangement, there is also a diagonal arrangement, which we point out here in particular (Box 1 in relation to Box 2). By this one can notice that the TMA corresponds with the Periodic table of chemical elements (PT), through the validity of both Mendeleev principles: the principle of continuity and the principle of minimum change. Since both these principles are also apply to the amino acid code (Swanson, 1984), all together support the hypothesis of a prebiotic complete amino acid code, i.e. genetic code (Rakočević, 2004).

The presented sequence also points to the logic of choice the uniqueness in the case of the selection of protein amino acids within the standard genetic code, minimal in two manner. So, in this paper, from the sequence (00-11-22-11-00), located in the starting column of Table 3, generated from the initial triplet of the last column (00-11-22) in the "Table of minimal adding" (TMA), explained in Remark 1, follows: $[00 + (2 \times 10) = \mathbf{20}]$, $[11 + (2 \times 10) = \mathbf{31}]$, $[22 + (2 \times 10) = \mathbf{42}]$, $[11 + (5 \times 10) = \mathbf{61}]$, $[00 + (5 \times 10) = \mathbf{50}]$. So, we get the sequence (**20-31-42-61-50**), analogical to the the number of atoms in five rows of 20 AAs (within their side chains) as in Table 2;

There is also an other logic of choice the uniqueness, presented in a previous paper (Rakočević, 2011b) as follows: $[(25-36-26-16) (16-17-18)] \rightarrow [(6-5 = 01) (36-25 = 11)] [(26), (26+16 = \mathbf{42}), (42+17 = \mathbf{59}), (59+18 = \mathbf{77})]$, we get the sequence (**26-42-59-77**), analogical to the number of atoms in four diversity types of 20 AAs (within their side chains), dispersed in four columns of a specific arrangement (Figure 3 in Rakočević, 2011b).

In addition to the above, there are other relationships with the amino acid (genetic) code. Within TMA there are two horizontally adjacent numbers, 5 and 6 ($6 - 5 = 01$) whose squares (25 and 36, respectively) are vertical neighbors ($36-25 = 11$); and this is a unique case, the only one in the whole TMA. Analogy (and correspondance) with the amino acid (genetic) code is in the sense that within the Table of genetic code (TGC) 25 codons are coded for (2+4) amino acids of lower complexity (possess C & H, or only H, in side chain), while 36 codons are coded for (6+8) amino acids of more complexity, which besides C & H, possess even more the atoms of other elements: N, O or S (cf. Table 6 in Rakočević, 2011b, p. 829). On the other hand, in the sequence $[(25-36-26-16) (16-17-18)]$ the sum of the three numbers with the status "to be added" is $16 + 17 + 18 = 51$, which is a fourth of the sum of the four numbers in the given sequence: $26 + 42 + 59 + 77 = 204 = 4 \times 51$. And this is the only such case in the TMA. For example, such a regularity does not apply to the TMA analogue sequence $[(36-47-37-27) (27-28-29)]$, because $27+28+29 = 84$ is not a fourth of the sum of the four numbers in an analogously obtained sequence: $37 + 64 + 92 + 121 = 314 \neq 4 \times 84$.

The "enigmatic" question why, in the arrangement of Table 3 (in stating column), the first *three* rows increase by 2 units, and the other *two* rows per 5 units,² it stops to be mysterious when we know that the ratio of numbers 2 and 5, in binary form (010/101), represents the first possible connection between the "golden route" on Farey tree and the path of the permanent change (0101010101 ...) on the binary tree of the genetic code (Rakočević, 1998).

Box 1. The logic of a parallel choice

The important feature of the "Table of Minimal Adding" (TMA), within the limits of the two-digit arrangement of decimal number system, is a specific diagonal structure, with 10 odd and 9 even of diagonals. Here, in Box 2, only odd diagonals are presented, because among them there is a diagonal (the last row in the second quadrant of Box 2), which analogically corresponds with a specific chemical distinct arrangement of protein amino acids (Fig. 9 in Rakočević, 2011b, p. 834; here: fourth quadrant in Box 2): in 8+8 AAs of alanine stereochemical type (within their side chains) there are 91 and 81 atoms, respectively; in 4+4 AAs of alanine stereochemical type plus 1+1 AAs of valine stereochemical type there are 71 and 61 atoms, respectively; in 4+4 AAs of alanine stereochemical type plus 1AA of glycine stereochemical type, plus 1 AA of proline stereochemical type there are 41 and 31 atoms, respectively. Finally, within the set of four AAs of non-alanine type, 2+2 AAs possess 21 and 11 atoms, respectively. [About four amino acid stereochemical types see in (Popov, 1989) and in (Rakočević & Jokić, 1996.)

Box 2. The arrangement followed from diagonal structure of TMA

(01) [1] (03,13,23) [3] (05, 15, 25, 35, 45) [5] (07, 17, 27, 37, 47, 57, 67) [7] (09, 19, 29, 39, 49, 59, 69, 79, 89) [9]	[1] (99) [3] (77, 87, 97) [5] (55, 65, 75, 85, 95) [7] (33, 43, 53, 63, 73, 83, 93) [9] (11, 21, 31, 41, 51 , 61, 71, 81, 91)																																				
11+ 91 = 102 (91 – 11 = 8 x 10) 21+ 81 = 102 (81 – 21 = 6 x 10) 31+ 71 = 102 (71 – 31 = 4 x 10) 41+ 61 = 102 (61 – 41 = 2 x 10) 51 / 51 (The top two quadrants possess five diagonals each in the first and second part of the TMA, respectively. All other explanations are in Box 1)	<table border="1" style="width: 100%; text-align: center;"> <tr> <td>S₀₅</td> <td>T₀₈</td> <td>L₁₃</td> <td>A₀₄</td> <td>G₀₁</td> <td>31</td> </tr> <tr> <td>D₀₇</td> <td>E₁₀</td> <td>M₁₁</td> <td>C₀₅</td> <td>P₀₈</td> <td>41</td> </tr> <tr> <td>K₁₅</td> <td>R₁₇</td> <td>Q₁₁</td> <td>N₀₈</td> <td>V₁₀</td> <td>61</td> </tr> <tr> <td>F₁₄</td> <td>Y₁₅</td> <td>W₁₈</td> <td>H₁₁</td> <td>I₁₃</td> <td>71</td> </tr> <tr> <td colspan="2">91</td> <td colspan="2">81</td> <td>GV</td> <td>11</td> </tr> <tr> <td colspan="2"></td> <td colspan="2"></td> <td>PI</td> <td>21</td> </tr> </table>	S ₀₅	T ₀₈	L ₁₃	A ₀₄	G ₀₁	31	D ₀₇	E ₁₀	M ₁₁	C ₀₅	P ₀₈	41	K ₁₅	R ₁₇	Q ₁₁	N ₀₈	V ₁₀	61	F ₁₄	Y ₁₅	W ₁₈	H ₁₁	I ₁₃	71	91		81		GV	11					PI	21
S ₀₅	T ₀₈	L ₁₃	A ₀₄	G ₀₁	31																																
D ₀₇	E ₁₀	M ₁₁	C ₀₅	P ₀₈	41																																
K ₁₅	R ₁₇	Q ₁₁	N ₀₈	V ₁₀	61																																
F ₁₄	Y ₁₅	W ₁₈	H ₁₁	I ₁₃	71																																
91		81		GV	11																																
				PI	21																																

At the end of this analysis of the relationship between the three arrangements, presented in the three tables (Tables 1-3), we present another more, perhaps slightly hidden, relationship. Namely, from the relationship between the first two and the last two rows of numbers in the two

² The answer to the question why 3 versus 2 rows is currently not known. However, I suppose that this could be due to the fact that "the limit of the golden numbers is 3/2" (Moore, 1994).

upper structures within Box 2, it follows that there is a correspondence between the Table of Minimal Adding (TMA) and the Multiplication Table (TM), as well as between their diagonal structures in the decimal number system. On the other hand, if TM is written adequately, systemic and systematically,³ as TMA, then, on both sides of the main diagonal, there are the numbers: 0, 2, 6, 12, 20, ... (Table 1.1 and Rakočević, 2011b), which represent the values for "q" in the equation of the generalized golden mean (Spinadel, 1998, 1999; Rakočević, 2004b, 2011b). It follows from this that it can be understood why the amino acid (genetic) code is determined by the golden mean (Rakočević, 1998, 2011b) (Remark 2).

Remark 2. The determination of the amino acid (genetic) code with the golden mean leads to the CIPS (Cyclic Invariant Periodic System), in which the positions of five classes of AAs are strictly determined, two in the less complex and three in the more complex superclass: 1. (SC-TM), 2. (GV-PI), 3. (DE-NQ), 4. (AL-KR), 5. (FY-HW). Less complex aliphatic AAs in the side chains, besides C & H (Gly only H), contain the less polar nitrogen atom (N), and no longer contain the polar atom of oxygen (O), while more complex AAs contain it. Aromatic AAs also fall into the superclass of more complex AAs. In addition, we recall that in all the presented AAs pairs, the first member, as a smaller molecule, belongs to class II of AAs, handled by class II of enzymes aminoacyl-tRNA synthetases, and the second one to class I. [Cf. Figs 6 & 7 in (Rakočević, 2011b, pp. 832-833) and 4th quadrant within Box 2 in this paper.]

However, if we note that the said numbers on the main diagonal of TM, taken in a double amount (0, 4, 12, 24, 40, ...), represent the first cathete in the Pythagorean–Diophantine triangle set, and increased by one – the hypotenuse of the same triangles: 1, 5, 13, 25, 41, ... [The second cathete is "taken" from the set of odd natural numbers, respectively.] Thus, it is possible to understand why the amino acid (genetic) code was determined by the first (3-4-5) Pythagorean–Diophantine triangle (Shcherbak, 1994). In addition, considering the fact that there is a "coherence of the chemical and genetic code" (Rakočević, 1991)⁴, more precisely, there is an analogous unity of the chemical and genetic code (Rakočević, 2017), that is, if everything here presented, taken together, then there can be enough understanding for the standpoint that "the genetic code is an intelligent design ..." (Castro-Chavez, 2010, p. 718).⁵

³ "Systemic and systematically", that is, correspondent with the principles of continuity and minimum change (which, as we said, apply to both – genetic code and chemical code; cf. footnote 4). Practically, this means writing first the sequence (0, 1, 2, ..., 9) as a vertical column, with the status of the first column. Then the results of the multiplication are written: the results of the multiplication by zero yield the zeroth column, to the left of the first column; the results of the multiplication by the unit give this same, start-up, t.e. the first column. Next, continuous columns of multiplication with 2, 3, etc

⁴ "... the chemical code, built on the very principles mentioned and in complete accordance with the genetic code. ... All the relations in the chemical code and the genetic code are in accordance with periodicity and cyclicity of the natural number system ..." (Rakočević, 1991, p. 1). (Cf. footnote 9.)

⁵ "With insight into the results ... one is forced to propose a hypothesis (for further researches) that here, there really is a kind of intelligent design; not the original intelligent design, dealing with the question – intelligent design or evolution, which is rightly criticized Here, there could be such an intelligent design, which we could call 'Spontaneous Intelligent Design' SPID) that is consistent with that design which was presented by F. Castro-Chavez (2010), and is also in accordance with the Darwinism." [Additional note: The evidence that Darwin tacitly advocates the idea of intelligent design, learns from the insight that his only illustration (Diagram) in *Origin of Species* is

We again look at Box 2 (quadrant fourth), from the aspect of the viewing of possible distinct chemical pairs of AAs. We see that the pairs strictly correspond to the pairs found in two classes of AAs handled by two classes of enzymes aminoacyl-tRNA synthetases (Wetzel, 1995; Rakočević, 1997a). We first consider the AAs in the non-alanine type set, and in addition, in the quantum "11" and the quantum "21"; the result of matching is obvious: a larger molecule in class I, while lower in class II, GV and PI, respectively. Then we look at the pairs in quantum "81" and "91". In all cases, except in one, in one column there are larger molecules (class I), and in the second one smaller molecule (class II). The exception is a pair CM in quantum "81" where both AAs belong to class I, and a pair of ST in quantum "91" where both AAs belong to that class II. (Cf. Survey 4 in Rakočević, 1998.)⁶

*

Not only in two decades (as in Table 1), the balance of the number of atoms is contained also within the halves of both decades, in four quintets: 1. GAVPI, 2. LKRQE, 3. NDSTC, 4. MFYWH. However, the said regularities are expressed (become visible) only when the four quintets develop into a system of five quartets; the five upper and five lower pairs are arranged according to the next logic: the first pair of the upper quintet with the first pair of the lower one, the second with the second, etc. (Table 4). If in fact the resulting system (with five quartets) perceived as a cyclic system (the first row in a cyclization occurs also as the last), then the balances (and symmetry) exhibit a specific self-similar arrangement more: two times appear the same quantities of the number of atoms, although with the different qualities (with different AAs) in four cases (117, 117) (118, 118), (119, 119), (120, 120). Such a repetition means *per se* a correspondence with the principle of self-similarity (Remark 3).

But, not only that, the pair 119-120 stand in relation to the same pair in the Rumer's Table of AAs and their corresponding nucleotide doublets (Table 5); The number of 118 atoms is also found as a half number of atoms within a "specific amino acid arrangement" (Rakočević & Jokić, 1996, Survey 1);⁷ and 117 atoms as the number of hydrogen atoms in all 20 protein AAs, within their side chains. [Also, number of hydrogen atoms within amino acid molecules appears to be in relation to natural numbers sequence; so, in side chains there are: 1, (2), 3, 4, 5, 6, 7, 8, 9, 10 of hydrogen atoms, and in whole molecules: 5, (6), 7, 8, 9, 10, 11, 12, 13, 14 (Sukhodolets, 1985) (Cf. last sentence in legend of Table 7 on page 830, in: Rakočević, 2011b).]

actually a strict mathematical program corresponding, directly and / or indirectly, to the structures currently known within TM & TMA (Rakočević, 1994, 2017).]

⁶ These pairs are exceptions also in the basic chemical characteristic. The CM is the only pair in which both members, apart from atoms of the first and second periods of PT of chemical elements, possess one atom from the third period. On the other hand, the ST and CM pairs are chemically similar and constitute a set of halcogen AAs, over oxygen and sulfur, both from the sixth group of halcogen elements. With this insight, it is easier to understand why they are at the center of CIPS: they start counting from them and have an ordinal number 1.

⁷ Specificity is, among other things, in the fact that AAs of alanine stereochemical type are taken by once and of non-alanine type (GP-VI) by twice (cf. Remark 3). In addition, the specificity is that "specific amino acid arrangement" exists in a direct connection with the system of amino acids precursors, with their order of inclusion, thus confirming each other - the choice of AAs and their precursors.

The principle of self-similarity extends, however, substantially broader within the amino acid (genetic) code. Moreover, this extension is present parallel with Shcherbak's principle of "analogies with quantum physics." [Details of this analogy can be seen in our previous work (Rakočević, 2017).] Thus, in the set of protein AAs, in their side chains, there is a pattern 117/87 (117 hydrogen atoms and 87 non-hydrogen atoms). This pattern corresponds to the pattern 107/97 where 107 is the number of all atoms in 10 polar AAs [(KR-ST-DE-NQ) + (YH)], and number 97 represents the number of atoms in 10 AAs that are not polar [(GP-AL-VI) + (CM-FW)] [The grouping of AAs follows from the distinction of AAs in four types of diversity (Rakočević, 2011a, 2011b).] In this classification calculated purely polar AAs only, and not those that are semi-polar (GP and W). Classification by the polarity criterion, however, follows from the polarity of the functional groups, that is, from the electronegativity of the atoms that make them. In view of the "chemical eyes", this polarity/nonpolarity is directly visible, and it agrees with the strict parameters for which this polarity is measured [polar requirement (Woese et al., 1966), hydrophathy index (Kyte & Doolittle, 1982) and cloister energy (Swanson, 1984)], as we have shown in previous works (Rakočević & Jokić, 1996; Rakočević, 2004a, Section 3.3).

A direct self-evident of polarity/nonpolarity follows from the facts as stated herein: aliphatic amino acids that in the side chains possess only electropositive atoms, C & H, or with them also possess an electropositive sulfur atom more, are nonpolar (AL-VI-CM). [Under electropositive we consider those atoms whose electronegativity (according to Pauling) is 2.5 or less, while values with more than 2.5, the atoms are electronegative (polar).] Aromatic amino acid phenylalanine (F), which in the side chain only has a nonpolar benzene ring, is also nonpolar amino acid. On the other hand, aliphatic amino acids that in the side chains, except carbon and hydrogen (C & H), also have electronegative nitrogen atoms (N) and/or oxygen (O), are polar amino acids (KR-NQ-ST-DE); aromatic amino acid tyrosine (Y) which in the side chain has a phenol group is also a polar amino acid; Finally, the aromatic amino acid histidine (H), which in the side chain has a polar imidazole ring, is also a polar amino acid.

On the third side, there are three more amino acids, that are not polar but semi-polar. First of them is glycine (G). Its molecule has two equivalent hydrogen atoms, which at the same time exist in both the "head" and the side chain of the amino acids. Although the hydrogen atom itself is electropositive (nonpolar), the influence of the polar "head", i.e. the influence of the polar amino acid functional group is inevitable, and hence the semi-polarity of this amino acid is followed. Similarly, the proline has three nonpolar CH₂ groups in the side chain, but one of the two ending allyl groups is attached to the polar nitrogen atom in the "head" of the amino acid; hence its semi-polarity follows. Aromatic tryptophan (W), which in the side chain contains an indole ring, composed of one nonpolar benzene and one polar pyrrole ring, must also be a semi-polar amino acid.

It is understood that it makes sense to ask a question about the classification of protein amino acids into the class of strictly nonpolar amino acids (7 AAs) and the class of polar amino acids

where semi-polar AAs also come (total 13 AAs). Such a classification shows a strict separation of the TGC into polar/nonpolar as well as inner/outer space with the differences for 1 molecule and 1 atom, respectively (Rakočević, 2000, Section 3, Eq. 1-4, pp. 279-280; 2011b, Eq. 11-14 on p. 838 and Tab. A.3 on p. 840).

Remark 3. So far, the chemical meaning for the duplicates of AAs in TGC has been shown as following: SLR (Crick, 1968); SLIR (Rakočević, 2007, Table 7; Wohlin, 2015, Table 2); GP-VI (Rakočević and Jokić, 1996, Survey 1); and here in this paper: GNLM in Table 4. All these duplications, *ipso facto*, support the hypothesis of a prebiotic complete genetic code (Rakočević, 2004), which is now known as the standard genetic code ("Canonical genetic code"), and all other genetic codes are "deviant codes" (Weaver, 2012, pp. 568-569), or exceptions from the standard genetic code. These codes, *per se*, do not change nothing on the fact, when it comes to the genetic code as "amino acid code" (Swanson, 1984), because they are represented in all variants with the same 20 protein amino acids. (The only exceptions are Pyrrolyisine and Selenocysteine, presented at some very few organisms.) [Knight at al., 2001, p. 49: "The genetic code evolved in two distinct phases. First, the 'canonical' code emerged before the last universal ancestor; subsequently, this code diverged in numerous nuclear and organelle lineages".]

*

All together in the question are systemic and systematic natural arrangements, whose organization and determination correspond with the principle of self-similarity.⁸ The already well-known facts that genetic code represents an analogy with natural (verbal) language are joined now to the facts about analogies between genetic code arrangements and specific arrangements within the set of natural numbers.⁹ Such an agreement (through "ideal" and "intelligent" structures and arrangements) leads us to the conclusion that the genetic code was indeed complete even in prebiotic conditions (as a set of 20 canonical amino acids and the set of 2+2 pyrimidine / purine canonical bases, respectively); and the "evolution" of the genetic code can only mean a degree of freedom of standard genetic code, i.e. the possible exceptions and deviations from the standard genetic code, as we have presented here and in previous works, and, as also other authors have pointed out (Sukhodoletc, 1981, 1985; Swanson, 1984; Rakočević, 1988a, 1988b, 2004a; Popov, 1989; Shcherbak, 2008; Castro-Chavez, 2010). And, secondly, there is no point in talking about the evolution of genetic code in terms of the evolution of organisms, but only about the degree of freedom of a unique coherent and harmonious system, which degree of freedom is expressed in the form of exceptions and/or deviations (Remark 3) .

All the present arrangements and regularities are an addition to the previously revealed systems and arrangements of protein amino acids (as constituents of GC) in which chemical distinctions and classifications are accompanied by arithmetical and algebraic regularities, both

⁸ "In correspondence with this, Complete Genetic Code must be based on several key principles. We are going to list only those considered to be the most important: 1. The principle of systemic self-related and self-similar organization. ..." (Rakočević, 2004, p. 231).

⁹ "In determination of the genetic code, except two inherent alphabets – twenty amino acids and four amino bases (two pyrimidines & two purines – is involved still one "hidden alphabet", the series of natural numbers, with all its regularities and laws" (Rakočević, 2011a, p. 4). (Cf. footnote 4.)

in our previous works and in the works of other authors (Rumer, 1966; Kyte & Doolittle,1982; Doolittle,1985; Sukhodolets,1985; Leunissen and De Jong,1986; Taylor & Coates, 1989; Koruga,1992; Shcherbak, 1994; Damjanovic, 1998; Dragovich and Dragovich, 2006; Négadi, 2009; Castro-Chavez, 2010; Mišić, 2011; Petoukhov, 2016; Dragovich et al, 2017). On the other hand, all these results together show that the genetic code is a deeper essence than would follow only from the chemical properties of the amino acid molecules. And it also means that *biological*, as such, is a deeper essence than it follows only from physics and chemistry.

Acknowledgement

I am grateful to Branko Dragovich, Vladimir Ajdačić and Nataša Mišić for support and helpful, stimulating discussion.

TABLES

Table 1. The chemically determined order of protein amino acids

(1) G	01	08	N	(11)
(2) A	04	07	D	(12)
(3) V	10	05	S	(13)
(4) P	08	08	T	(14)
(5) I	13	05	C	(15)
(6) L	13	11	M	(16)
(7) K	15	14	F	(17)
(8) R	17	15	Y	(18)
(9) Q	11	18	W	(19)
(10) E	10	11	H	(20)
	<hr/>	<hr/>		
	102	102		
o/e	51±1	51±1		

The 20 protein AAs, arranged into two decades in accordance to ordinal amino acid number, 1-10 and 11-20; the numbers presented outer: the ordinal numbers 1-20; the numbers presented inner: the number of atoms within side chain of the responding amino acid. The designations "o/e" as "odd / even". For details see the text.

Table 2. The order of five quartets of protein amino acids

G (01)	A (04)	N (08)	D (07)	→	20
V (10)	P (08)	S (05)	T (08)	→	31
I (13)	L (13)	C (05)	M (11)	→	42
K (15)	R (17)	F (14)	Y (15)	→	61
Q (11)	E (10)	W (18)	H (11)	→	50
51-1	51+1	51-1	51+1		

The explanation in the text.

Table 3. A specific natural numbers arrangement

00	02	04	06	08	10	12
11	13	15	17	19	21	23
22	24	26	28	30	32	34
11	16	21	26	31	36	41
00	05	10	15	20	25	30
44	60	76	92	108	<u>124</u>	140
	12	14	16	18	20	22
	23	25	27	29	31	33
	34	36	38	40	42	44
	41	46	51	56	61	66
	30	35	40	45	50	55
	140	156	172	188	<u>204</u>	<u>220</u>
	22	24	26	28	30	32
	33	35	37	39	41	43
	44	46	48	50	52	54
	66	71	76	81	86	91
	55	60	65	70	75	80
	220	236	252	268	<u>284</u>	<u>300</u>
	32	34	36	38	40	42
	43	45	47	49	51	53
	54	56	58	60	62	64
	91	96	101	106	111	116
	80	85	90	95	100	105
	300	316	332	348	364	380
	...					

The arrangement represents the Table of distinct 2-5 adding (TDA) with starting column which follows from TMA, explained in Remark 1. All other explanations in the text.

Table 4. A specific protein amino acids arrangement

119					
G 01	N 08	L 13	M 11	(33)	120
A 04	D 07	K 15	F 14	(40)	
V 10	S 05	R 17	Y 15	(47)	117
P 08	T 08	Q 11	W 18	(45)	
I 13	C 05	E 10	H 11	(39)	
G 01	N 08	L 13	M 11	(33)	
24/13	18/23	40/39	37/43	118/119	
(37)	(41)	(79)	(80)	117/120	
118					

The first row is repeated at the bottom, and thus one cyclic system is obtained. There are 117 atoms in two outer columns; at even positions 118, at odd 119; in two inner columns 120 atoms. On the other hand, in the lower half of the Table there are 117 atoms ones more; in the lower diagonally "wrapped" area 118, and in the upper 119; in the upper half of Table 120 atoms. The repeated four AAs at the bottom of the Table make to achieve a diagonal balance with a difference of only one atom; moreover, to establish a sequence from the series of natural numbers: 117, 118, 119, 120. (About generating the Table see in the text.)

Table 5. The Rumer's classification of amino acids into two classes (Rumer, 1966)

114	30 116	(119)	89 108	125
Gly	GG (6)	Phe	UU (4)	Leu
Pro	CC (6)	Asn	AA (4)	Lys
Arg	CG (6)	Ile	AU (4)	Met
Ala	GC (6)	Tyr	UA (4)	ct
Thr	AC (5)	His	CA (5)	Gln
Val	GU (5)	Cys	UG (5)	Trp
Ser	UC (5)	Asp	GA (5)	Glu
Leu	CU (5)	Ser	AG (5)	Arg
125	36 106	(120)	84 118	114
330-66		330±00		
125 + 114 = 239 125 – 114 = 11				

The four-codon amino acids are on the left and the non-four-codon amino acids on the right. Each of two classes is classified into two subclasses, corresponding to the number of hydrogen bonds in the nucleotide doublets. On the other hand, nucleon number within amino acid molecules in these two classes is determined by Pythagorean triangle (Figure 1 in Shcherbak, 1994). At the bottom (shaded) – the number of atoms in the amino acid molecules (side chains): within 32 amino acid molecules on the left and 29 on the right, within the set of "61" amino acid molecules, each molecule encoded by one codon. [There are the same results, 330-66 (as 8 x 33) and 330 (as 10 x 33) in standard GCT, both in relation with the result (9 x 33)±1 as we have shown in a previous work (Rakočević, 2004a, Table 3a on p. 224).] Up/down: 119/120 atoms in 11/12 amino acids (side chains), respectively, within the set of "23" amino acids. [Calculations: (30+89 = 119); (36+84 = 120); (30+84 = 114); (36+89 = 125); (125 – 114 = 11).]; [(05 x 6 = 30±0); (15 x 6 = 89+1); (06 x 6 = 36±0); (14 x 6 = 84±0)]; Up/down: (116/108 // 106 / 118) as the number of atoms within nucleotide doublets. (Note: In original Rumer's Table only the number of hydrogen bonds is calculated; all other calculations are ours.)

REFERENCES

- Castro-Chavez, F. (2010) The rules of variation: amino acid exchange according to the rotating circular genetic code, *J. Theor. Biol.* 264, 711-721.
- Crick, C.H. F. (1968) The Origin of the Genetic Code. *J. Mol. Biol.* 38, 367-379.
- Damjanović, Z., (1998) Logic core of genetic code, *Glasnik Sect. Nat. Sci. Montenegrin Acad. Sci. Art (CANU)* 12, 5-8.
- Damjanović, Z. M., Rakočević, M. M. (2005) Genetic code: an alternative model of translation. *Annals of New York Academy of Sciences*, 1048, 517-523.
- Doolittle, R.F. (1985) Proteins, *Scientific American*, 253, 74-85.
- Dragovich, B., Dragovich, A. (2006) p-Adic Model of DNA Sequence and Genetic Code, arXiv:q-bio/0607018v1 [q-bio.GN].
- Dragovich, B., Khrennikov A. Yu., Mišić N. Ž. (2017) Ultrametrics in the genetic code and the genome. *Applied Mathematics and Computation* 309, 350-358. arXiv:1704.04194 [q-bio.OT].
- Kyte, J., Doolittle, R. F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157, 105-132.
- Knight, R.D., Freeland, S.J., Landweber, L.F. (2001) Rewiring the keyboard: evolvability of the genetic code, *Nature Reviews Genetics*, Jan., 2001, Vol. 2, no 1, pp. 49-58.
- Koruga, D.L. (1992) Neuromolecular Computing. *Nanobiology*, 1, 5-24.
- Leunissen, J. A., De Jong, W.W. (1986) Phylogenetic trees constructed from hydrophobicity values of protein sequences. *J. Theor. Biol.* 119, 187-196.
- Mišić, N.Ž. (2011) Nested numeric/geometric/arithmetic properties of shCherbak's prime quantum 037 as a base of (biological) coding/computing, *Neuroquantology* 9(4), 702-715.
- Moore, G.A. (1994) The limit of the golden numbers is $3/2$. *The Fibonacci Quarterly*, June-July, 211-217.
- Négadi, T. (2009) The genetic code degeneracy and the amino acids chemical composition are connected, *Neuroquantology*, Vol. 7, 1, 181-187; arXiv:0903.4131v1 [q-bio.OT].
- Petoukhov, S. (2016) The system-resonance approach in modeling genetic structures, *BioSystems*, 139, 1-11.
- Popov, E. M. (1989) *Strukturnaya organizaciya belkov*. Nauka, Moscow (in Russian).
- Rakočević, M. M. (1988a) Three-dimensional model of the genetic code, *Acta Biologiae et Medicinae Experimentalis*, (Prishtina), 13, 109-116 [An excerpt in: <http://www.rakocevcode.rs>.]
- Rakočević, M. M. (1988b) Genes, Molecules, Language (in Serbian with an English Language Supplement), Naučna knjiga, Belgrade. (<http://www.rakocevcode.rs>)

- Rakočević, M.M. (1991) The coherence of the chemical and genetic code, *Proceedings of Faculty of science* (former: Faculty of philosophy), Chemistry Section, 2, 1-29, Niš.
- Rakočević, Miloje (Mirkov) (1994) *Logic of the Genetic Code*, Naučna knjiga, Belgrade. (<http://www.rakocevcode.rs>)
- Rakočević, M. M. (1997a) Two classes of the amino acyl-tRNA synthetases in correspondence with the Codon path cube. *Bull. Math. Biol.* 59, 645-648.
- Rakočević, M.M. (1997b) *The genetic code as a unique system*, Studentski kulturni centar, Niš (www.rakocevcode.rs)
- Rakočević, M.M. (1998) The genetic code as a Golden mean determined system, *Biosystems*, 46, 283-291.
- Rakočević, M. M. (2000) The factors of the classification of protein amino acids, *Proceedings, (Glasnik) of the Section of Natural Sciences of Montenegrin Academy of Sciences and art (CANU)*, 13, 273-294. (arXiv:q-bio/0611004 [q-bio.BM])arXiv:q-bio/0703011v2 [q-bio.OT].
- Rakočević, M.M. (2004a) A harmonic structure of the genetic code, *Journal of Theoretical Biology*, 229, 221– 234.
- Rakočević, M.M. (2004b) Further Generalization of Golden Mean in Relation to Euler's "Divine" Equation, *FME Transactions* (FME = Faculty of Mechanical Engineering, Belgrade, Serbia), 32, 95–98 (also in: arXiv:math/0611095 [math.GM]).
- Rakočević, M. M. (2006) Genetic code as a harmonic system, arXiv:q-bio/0610044v1 [q-bio.OT].
- Rakočević, M. M. (2007) A new genetic code table, arXiv:qbio/0703012v1 [q-bio.GN].
- Rakočević, M.M. (2011a) Genetic Code: Four Diversity Types of Protein Amino Acids, arXiv:1107.1998v2 [q-bio.OT].
- Rakočević, M.M. (2011b) Genetic code as a coherent system. *Neuroquantology* 9 (4), 821–841.
- Rakočević, M. M. (2013) Harmonic mean as a determinant of the genetic code, arXiv:1305.5103v4 [q-bio.OT].
- Rakočević, M. M. (2015) Enigma of Darwin Diagram. www.rakocevcode.rs submitted on 2015-01-06.
- Rakočević, M. M. (2017) Analogies of Genetic and Chemical Code. www.rakocevcode.rs (submitted on 2017-08-09 and new, minimally modified, version 2017-10-02).
- Rakočević, M.M., Jokić, A. (1996) Four stereochemical types of protein amino acids: synchronic determination with chemical characteristics, atom and nucleon number. *J. Theor. Biol.* 183, 345–349.
- Rumer, Yu, B. (1966) O sistematizaciji kodonov v genetičeskome kode, *Doklady Akad. Nauk. SSSR*, 167, 1393-1394.
- Shcherbak, V.I. (1994) Sixty-four triplets and 20 canonical amino acids of the genetic code: the arithmetical regularities. Part II, *J. Theor. Biol.*, 166, 475-477.

- Shcherbak, V. I. (2008) The arithmetical origin of the genetic code, in: M. Barbieri (ed.), *The codes of life: the rules of macroevolution* (pp. 153-181), Springer, Berlin.
- Spinadel, V.W. de (1998) The Metallic Means and design, In: *NEXUS II- Architecture and Mathematics*, Ed: Kim Williams.
- Spinadel, V.W. de (1999) The family of metallic means, *Visual Mathematics*, 1 (3), 1-16, <http://members.tripod.com/vismath1/spinadel/>.
- Sukhodolets, V. V. (1981) Evolucionie preobrazovaniya geneticheskogo koda, predskazivanie iskhodya iz gipotezy o fizicheskom predopredelenii naborov osnovaniy v sostave kodonov, *Genetika*, XVIII, 3, 499-501.
- Sukhodolec, V.V. (1985) The meaning of the genetic code: the reconstruction of the stages of prebiological evolution (in Russian), *Генетика*, XXI, 10, 1589 – 1599.]
- Swanson, R. (1984) A unifying concept for the amino acid code, *Bull. Math. Biol.* 46, 187-207.
- Taylor, R.J.F., Coates, D. (1989) The code within codons. *Biosystems* 22, 177–187.
- Weaver, R.F. (2012) *Molecular Biology*, fifth edition, McGraw-Hill, New York.
- Wetzel, R. (1995) Evolution of the Aminoacyl-tRNA Synthetases and the Origin of the Genetic Code. *J. Mol. Evol.*, 40, 545-550.
- Woese, C.R., et al. (1966) On the fundamental nature and evolution of the genetic code. In: Cold Spring Harbor Symp. *Quant. Biol.*, 31, 723-736.
- Wohlin, Åsa (2015) Numeral series hidden in the distribution of atomic mass of amino acids to codon domains in the genetic code, *J. Theor. Biol.* 369, 95–109.