ЦРНОГОРСКА АКАДЕМИЈА НАУКА И УМЈЕТНОСТИ
ГЛАСНИК ОДЈЕЉЕЊА ПРИРОДНИХ НАУКА, 15, 2003.
ЧЕРНОГОРСКАЯ АКАДЕМИЯ НАУК И ИСКУССТВ
ГЛАСНИК ОТДЕЛЕНИЯ ЕСТЕСТВЕННЫХ НАУК, 15, 2003.
THE MONTENEGRIN ACADEMY OF SCIENCES AND ARTS
GLASNIK OF THE SECTION OF NATURAL SCIENCES, 15, 2003.
UDK 575.113.1

Miloje M. Rakočević *

AN INSIGHT INTO HARMONIC STRUCTURE OF THE GENETIC CODE

A b s t r a c t

In this paper is presented a very harmonic structure of the genetic code within a system of "5 x 4" as well as of "4 x 5" of amino acids (AAs). In first case the five rows within the system start with one polar charged amino acid each, making first column, consisting from five polar charged AAs (D, R, K, H, E). Five polar non-charged AAs follow (N, P, Y, W, Q), then five non-polar AAs as last column (A, L, F, V, I), and finally, five polar or non-polar AAs, in a combination, as first to last column (A as non-polar; S, T as polar, and G & P as ambivalent AAs). A second system is subsequent to this one - "4 x 5" system with five nitrogen AAs (K, R, P, H, W), five oxygen (D, E, Y, S, T), five solely carbon (A, L, F, V, I) and five "combined" AAs (G with hydrogen as side chain; C and M with Carbon and Sulfur; N and Q with Carbon, Oxygen and Nitrogen). A strict balance of atom and nucleon number follows the classification in both systems.

*Prof. dr Miloje M. Rakočević, Department of Chemistry, Faculty of Natural and Mathematical Sciences, University of Niš, Višegradska 33, 18000 Niš, Serbia and Montenegro.

K e y w o r d s: Genetic code, canonical amino acids, harmonic structure, polarity, molecule mass, isotopes, nuclides, nucleotides, codons, codon rules, Siemion-Siemion's rule, Davidov's rule, Damjanović's spiral model, atom number, nucleon number, atom number balance, nucleon number balance.

SAGLEDAVANJE HARMONIJSKE STRUKTURE GENETSKOG KODA

I z v o d

U radu je predočena harmonijska struktura genetskog koda, u formi sistema od "5x4" i sistema od "4 x 5" aminokiselina (AAs), pri čemu je u prvom slučaju reč o pet redova (sa četiri kolone), a u drugom slučaju o četiri reda (sa pet kolona). Svaki od pet redova u sistemuu "5 x 4" započinje sa po jednom od polarnih naelektrisanih AAs, čime se realizuje i prva kolona (D, R, K, H, E). Sledeću kolonu čine pet polarnih nenaelektrisanih (N, P, Y, W, Q), poslednju pet nepolarnih (L, I, M, C, V) i pretposlednju - pet AAs od kojih je jedna nepolarna (A), dve polarne (S, T) i dve ambivalentne (G, P). S druge strane, prvi od četiri reda u sistemu "4 x 5" čine AAs koje u bočnim nizovima svojih molekula sadrže osim ugljenika još i azot (K, R, P, H, W), dok su u poslednjem AAs sa kiseonikom umesto azota (D, E, Y, S, T). U drugom redu nalaze se AAs sa ugljeničnim bočnim nizovima (A, L, F, V, I), i u pretposlednjem je jedna kombinacija AAs (G, sa vodonikom kao bočnim nizom; C i M koje uz ugljenik poseduju još i sumpor, i N i Q, koje uz ugljenik poseduju još azot i kiseonik). I jedna i druga klasifikacija praćene su ravnotežama broja atoma i nukleona u ovako dobijenim klasama i potklasama.

1. INTRODUCTION

In a previous paper (Rakočević, 1998) we presented a system of "2 x 10" (10 pairs) of canonical amino acids (AAs), with firmly determined positions in pairs, within two classes, handled by class I and class II of enzymes aminoacyl-tRNA synthetases (Figure 1, in relation to

Figure 2)[1]. However, in this paper we will present a system of "5 x 4" (five quartets and four quintets at the same time) of AAs, with characteristics of a harmonic structure, "hidden" i.e. integrated within the system of "2 x 10" (Table 1). The system is being generated in

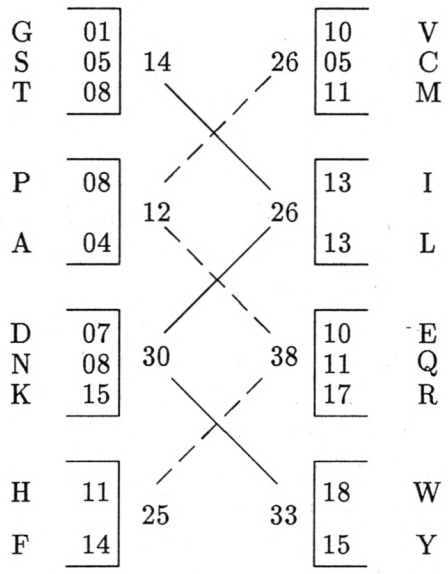


Figure 1. The 10 amino acids pairs in correlation to two classes of enzymes aminoacyl-tRNA synthetases (after Survey 4 in Rakočević, 1998, p. 290). The class II of enzymes aminoacyl-tRNA synthetases handles the smaller AAs within the pairs (on the left), whereas the larger (on the right) are handled by class I. The system is determined with a balanced proportionality of atom number in the same time. On the first (full) zigzag line there are 102+1 whereas on the second (dotted) line 102-1 atoms. Arithmetic mean for both 102±1. On the other side, within AAs (side chains) in left column there are 81, whereas in the right column 0123 of atoms. Notice that "81" (as 9 x 9) is the first possible (zeroth) arithmetic square in module 9, and 0-1-2-3 is the first possible (zeroth) logical square (as 00-01-10-11).

[1]The said ten AAs pairs are determined not only by amino acid - enzyme reactivity, but by another very important physical-chemical factors, such as hydrophilicity / hydrophobicity, polarity, acidity-basicity, volume and molecule mass, etc. (Rakočević, 2000).

three steps with the respect of the sequence of AAs within the system of "2 x 10" in the following solutions (Solutions 1-3):

$$\begin{matrix} D-E-R-K-H/W-Y-F \\ N-Q-F-Y-W \end{matrix} \quad (1)$$

$$\begin{matrix} D-R-K-H-E \\ N-F-Y-W-Q \end{matrix} \quad (2)$$

$$\begin{matrix} A-P-T-S-G \\ L-I-M-C-V \end{matrix} \quad (3)$$

Starting from aspartic acid in the step (1), remaining four polar charged AAs are being added successively, as a linear and continual series. Three remaining aromatic AAs are being added on these within one cyclic closure. Finally, the cycle is being closed by adding of two amide derivatives to their "mothers", on the same way they stand within the system of "2 x 10". The neighborhood of D-E within the step (2) is being realized over cycle: The pair D-N remains in the first, and the pair E-Q transfers in the last row. This way, the first and second column within the system of "5 x 4" are being realized, shown in Table 1. Finally, step three presents the manner on which the third and fourth column are being generated by taking two "pillars" from the system of "2 x 10" (Figure 1), added over the pair D-E.

A	T	L	M	W
P	E	Q	R	Y
I	D	N	K	F
G	S	V	C	H

Figure 2. The 10 amino acids pairs as "Natural system of amino acids" (simplified after Figure 42 in Dlyasin, 1998, pp. 68-69). The 8 pairs are identical as in our system in Fig 1, except two: G-A and V-L, instead G-V and A-L. From Table 1 it is clear that both solutions are valid (A-L and G-V as horizontal pairs and G-A and V-L as vertical).

2. A HARMONIC STRUCTURE

If we consider "harmony" as "a state of order, agreement, and/or completeness in the relations of things, or of parts of a whole to each other", then the harmonic structure of the system of "5 x 4" was realized on several ways and from several aspects: through the number of nucleons (Table 1) and molecule mass (Table 2); through third-letter codon rule (Siemion, Siemion, 1994) and first-third-letter codon rule (Tables 3, 3.1 and 3.2); then through first- second-letter codon rule (Davydov, 1998) (Table 4); through polarity and atom number (Table 4); finally, through nucleotides and codons balances and regularities (Tables 5 & 6).

					a	b	c	d	M
D	N	A	L	→	**189**	**189**	**221**	**221+3**	485.49 ≈ **485**
R	F	P	I	→	*289*	*289*	*341*	*341+0*	*585.70 ≈ 586*
K	Y	T	M	→	299	299	351	351+2	595.71 ≈ 596
H	W	S	C	→	*289*	*289*	*331*	*331+1*	*585.64 ≈ 586*
E	Q	G	V	→	**189**	**189**	**221**	**221+3**	485.50 ≈ **485**
					1255	1255	1465	1465+9	2738.04

Tab. 1. *The harmonic structure with two "acidic" and three "basic" amino acid quartets*

Four choices after four types of isotopes: (a) The number of nucleons within 20 AAs side chains, calculated from the first, the lightest nuclide (H-1, C-12, N-14, O-16, S-32); (b) The number of nucleons within 20 AAs side chains, calculated from the nuclide with the most abundance in the nature [the same patterns as in (a): H-1, C-12, N-14, O-16, S-32]; (c) The number of nucleons within 20 AAs side chains, calculated from the nuclide with the less abundance in the nature (H-2, C-13, N-15, O-17, S-36); (d) The number of nucleons within 20 AAs side chains, calculated from the last, the heaviest nuclide (H-2, C-13, N-15, O-18, S-36); (M) The AAs molecule mass. Notice that (d) is greater from (c) for exactly one modular cycle (in module 9) and that total molecules mass is equal 2 x 37^2. Notice also that molecule mass within five rows is realized through the same logic-patterns of notations as the first nuclide, i.e. isotope.

The sums of the number of nucleons within five horizontal rows i.e. quartets (as well as molecule mass in the column M) are presented in columns a, b, c, d of Table 1. In the first and last row ("acidic" quartets), because every upper differs from every lower AAs exactly for $\pm CH_2$ group of atoms, the number of particles (atoms and nucleons), as well as molecule mass, are equal. In three central rows ("basic" quartets) the sums of the number of particles, as well as the molecule mass, are in symmetry relation, realized through the validity of principle of continuity as well as of minimum change. (Cf. the three positions in the notations of the number of nucleons)[2].

D 133.10	N 132.12	A 89.09	L 131.18	→	485.49
R 174.20	F 165.19	P 115.13	I 131.18	→	585.70
K 146.19	Y 181.19	T 119.12	M 149.21	→	595.71
H 155.16	W 204.10	S 105.09	C 121.16	→	585.64
E 147.13	Q 146.15	G 75.07	V 117.15	→	485.50
755.78	828.88	503.50	649.88		2738.04
					2(37x37)

2 x 666
(36x37)

2 x 703
(38x37)

Tab. 2. *The distribution of molecule mass into four quintets and five quartets*
Explanation in the text.

[2]In the side chains of eight AAs within two "acidic" quartets there is 189+189 = 378 nucleons. On the other hand, the number 378 represents the sum of all numbers 1-27 in Shcherbak's modular Table of "multiples of 37" (in module 9), which determine the nucleon number within two classes: four-codon and non-four-codon AAs (Shcherbak, 1994, Table 1). (Notice also that 289 is equal 17^2). The continuity principle and the minimum change principle are also valid for the other genetic code structures, such as "Codon ring", "Mutation ring", "Codon path cube" and "Physical properties plot" in Swanson's work (Swanson, 1984, p. 187: "These relationships express the minimum change principle upon which the code appears to be founded").

Table 2 shows that this symmetry for molecule mass is characteristic not only for rows but also for columns, applying a strict determination with the Shcherbak's "Prime Quantum 037" (Shcherbak, 1994). If we compare the structure of the system of "5 x 4" in Table 2 with the codon structures in Figure 3 & 4, the regularities are self-evident. Molecule mass of AAs in two outer columns of Table 2 and the number of nucleons in AAs coded by two codon classes in Figure 4 are in a proportion relation of 703:703 = 1:1, while two inner columns in Table 2 and two classes in Figure 3 (codons of Py & Pu type) are in proportion relation of 666: 999 = 2:3. [*Remark 1*: The Harmonic mean (H) of a whole (a) and its half (b) equals exactly 2/3 of the whole, H(a, b) = 2ab / (a+b)].

(a). Triplets with two identical pyrimidines

TTC	TTA	TTG	CCT	CCA	CCG
CTT	ATT	GTT	TCC	ACC	GCC
TCT	TAT	TGT	CTC	CAC	CGC
F91	L57	L57	P42–1	P42–1	P42–1
L57	I57	V43	S31	T45	A15
S31	Y107	C47	L57	H81	R100

Total Nucleon Number of the Side Chains

999

(b). Triplets with two identical purines

AAT	AAC	AAG	GGT	GGC	GGA
TAA	CAA	GAA	TGG	CGG	AGG
ATA	ACA	AGA	GTG	GCG	GAG
	333			333	
N58	N58	K72	G1	G1	G1
*Stop*0	Q72	E73	W130	R100	R100
I57	T45	R100	V43	A15	E73
		333			

Total Nucleon Number of the Side Chains

999

Figure 3. The triplets with two identical bases and unique one (after Table 1 in Shcherbak, 1996). This Figure shows the balances of nucleon sums and peculiarity of their writing in the decimal system. The central axis corresponds to Rumer's rules (Rumer, 1966) for TCAG → GACT or Triplet → Antitriplet transformations.

The structure of the system of "5 x 4" with (multiplied) quantum of molecule mass (**36** x 37), (**37** x 37) and (**38** x 37), as it was notified in Table 2, corresponds with two more Shcherbak's classes, with four-codon and non-four-codon amino acids (Shcherbak, 1994, Figure 1),

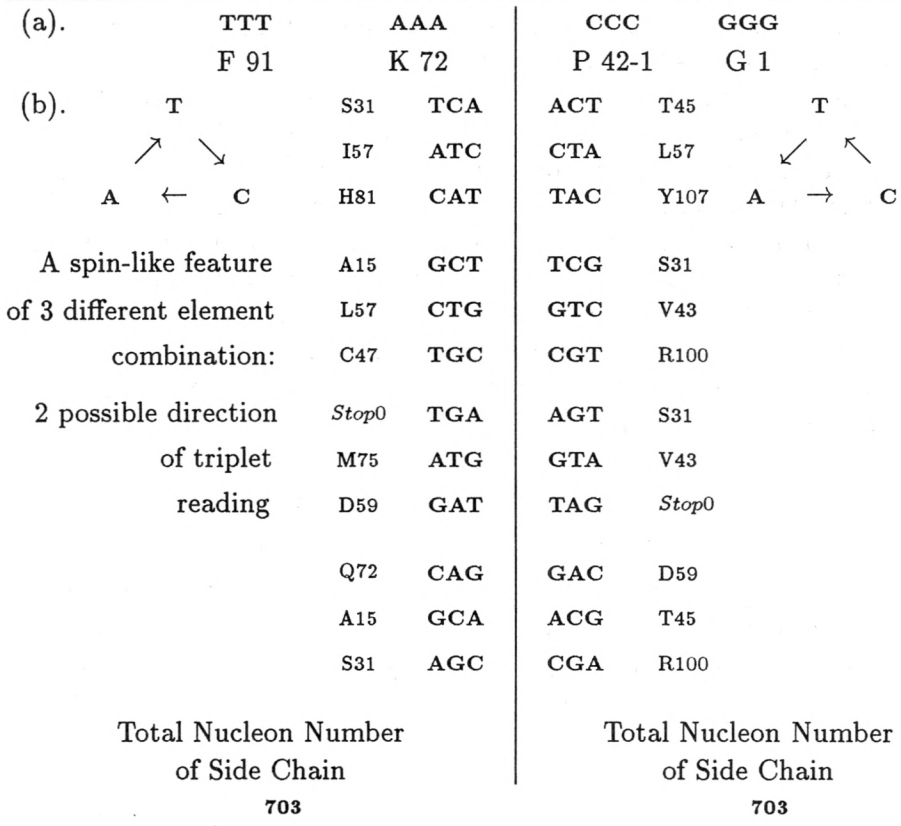


Figure 4. The triplets with three identical (a) and three unique bases (b) (after Table 3 in Shcherbak, 1996). The Figure shows the balance of nucleon sums by one of two relative positions of parts (a) and (b). The axis corresponds to Rumer's (Rumer, 1966) and Spin → Antispin transformations. The Triplet → Antitriplet transformation keeps their both positions or the one and the same side relative to the axis (an example of triplet-antitriplet pair: TCA, TGA). Notice four triplet group within the subsystem of three unique bases (b), each group with six permutations: (1) TCA without G, (2) GCT without A, (3) TGA without C and (4) CAG without T. After our hypothesis (for further researches) the missing base in form of codon with three identical bases (GGG, AAA, CCC, TTT, respectively) appears to make a "syn-codon" for all of four groups: GGG/TCA, AAA/GCT, CCC/TGA and TTT/CAG. It is clear that in RNA- code instead T goes U.

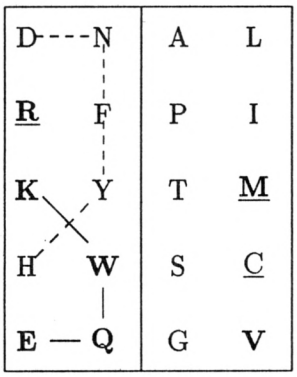
where there are in total (within side chains) 39 x 37 nucleons (333+1110 = **39** x 37). [Note: Koruga is shown that the number sequence 36, 37, 38, in correspondence with the sequence 11, 12, 13, is very immportant in determination of the symmetry of microtubule proteins (Koruga & Krstic, 1990)].

3. CODON STRUCTURE RULES

Table 3 shows relations with third-letter codon position, i.e. with Siemion-Siemion's rule (Siemion, Siemion, 1994), which appears to be in a strict accordance to the Damjanović's spiral model of the genetic code (Damjanović, 1998; Rakočević, 2002; Damjanović, 2003: personal communication), as it is shown in next solution (Solution 4), where the numbers are given in quaternary numbering system, and where the sign θ means an "empty space", and the sign * a "stop comand"[3]:

K(00)-Q(01)-E(02)-*(03)-T(10)-P(11)-A(12)-S(13)- θ(20) -[...] - *F(33)*
***F-N(00)-H(01)-D(02)-Y(03)*-W(10)-M(11)** (4)
(100) -(101) - (102) - (103)- (110) - (111)

θ(20) - [R(21)-G(22)-*C(23)*-I(30)-L(31)-V(32)-] - *F(33)*



Tab. 3. *The distribution of AAs in harmonic structure after codon rules*
Within two columns on the right there are "complex" AAs (except two sulfur AAs, which are "simple": Cysteine of Py type and Methionine of Pu type); complex, because they are coded by codons that have either pyrimidine, or purine in the third codon position. On the left are "simple" AAs (except arginine, which is "complex") - italic of Py type and bold of Pu type (cf. Tables 3.1 - 3.2).

[3]Three "stop codons" are three derivatives of (four-letter) amino bases, i.e. two and two Py- Pu alphabet. On the other hand, one "stop command" is one more "letter" within amino acid alphabet.

3.1. The Siemion-Siemion's rule

The Siemion-Siemion's rule relates to difference of the bases in third position within the codon. [Siemion, Siemion, 1994, p. 139: "It is shown that in the pairs of amino acids coded by the codons possessing identical bases in the first and second positions, the amino acids with R in the third position are of higher structural importance ..., than the amino acids coded with Y. The same structural factors seem to be of importance for the codon choice in the case of amino acids coded by more than two codons. The amino acids which prefer alpha- helical over the beta-sheet conformation favour the codons with R in their third position, and those which favour the beta-sheet conformation favour the codons with Y in this position."][Note: From Koruga's work (Koruga, Tomič, Ratkaj, 2002) it follows that the relation between second and third base within the codon is responsible for random protein conformation]. From all Siemion-Siemion's patterns (NNY & NNR) it follows that all 20 AAs can be classify into three classes: two

I - III	U	C	A	G	
Py - Py	54	26	52	44	176
Py - Pu	52	26	22	52	152
Pu - Py	46	24	30	12	112
Pu - Pu	44	24	50	36	154
	196	100	154	144	
	[(330–033) – 1]		**[(330–033)+1]**		

(330 – 066)±0 **330±0**

Tab. 3.1. *The atom number balances within Table of genetic code after first-third-letter codon position rule*
The number of atoms within AAs (side chains). First column (I-III) designates the type of the base in first-third position of corresponding codons. The letters U, C, A, G are related to four columns in Genetic Code Table. Within two inner and two outer rows as well as within two first and two second columns there are (8 x 33), [(9 x 33)±1], and (10 x 33) of atoms, respectively.

[(Py-Py)U]: $F_2(28)+L_2(26)=54$	[(Py-Pu)U]: $L_2(26)+L_2(26)=52$
[(Py-Py)C]: $S_2(10)+P_2(16)=26$	[(Py-Pu)C]: $S_2(10)+P_2(16)=26$
[(Py-Py)A]: $Y_2(30)+H_2(22)=52$	[(Py-Pu)A]: $^{**}(00)+Q_2(22)=22$
[(Py-Py)G]: $C_2(10)+R_2(34)=44$	[(Py-Pu)G]: $^{*}W(18)+R_2(34)=52$
[(Pu-Py)U]: $I_2(26)+V_2(20)=46$	[(Pu-Pu)U]: $IM(24)+V_2(20)=44$
[(Pu-Py)C]: $T_2(16)+A_2(08)=24$	[(Pu-Pu)C]: $T_2(16)+A_2(08)=24$
[(Pu-Py)A]: $N_2(16)+D_2(14)=30$	[(Pu-Pu)A]: $K_2(30)+E_2(20)=50$
[(Pu-Py)G]: $S_2(10)+G_2(02)=12$	[(Pu-Pu)G]: $R_2(34)+G_2(02)=36$

Simple combinations	*Complex combinations*	
[(Py-Py)]: F,Y,H,C (60)	[(Py-Py)] or [(Py-Pu)]: L,P	
[(Pu-Py)]: N,D	[(Pu-Py)] or [(Pu-Pu)]: I,T,V,A, G	(78+1)
[(Py-Pu)]:**Q,W** (66-1)	[(Py-Py)] or [(Py-Pu)] or [(Pu-Py)]: S	
[(Pu-Pu)]:**M,K,E**	[(Py-Py)] or [(Py-Pu)] or [(Pu-Pu)]: R	

Tab. 3.2. *The distribution of AAs in correspondence to the first-third-letter codon rule*

The 60 atoms within all AAs of Py type (encoded by codons with Py in third position) and 66-1 atoms within all AAs of Pu type, both in column "Simple combinations". Then, the 78+1 atoms within all AAs existing in column "Complex combinations"; (cf. the patterns of 60, 66 and 78 atoms within seven "golden" AAs, seven their complements and six non-complements, respectively, in Rakočević, 1998, Survey 2.1 and last passage on p. 289). On the other hand, inner space (in relation to dotted line) as well as outer space possess 102 of atoms each (within AAs side chains), what means an ideal balance. The sense of such a division by dotted line comes from the next logic. The AAs class "Py or Pu" type (first two rows) can be split into two subclasses: on the left there are AAs (L, P, I, T) with codons which possess Uracil or Cytosine in second codon position (two bases - U & C, one position - II); on the right there are AAs (V, A, G) encoded by codons which possess Guanine in first or second position (one base - G, two positions - I & II).

with "simple" and one with "complex" AAs. First class of "simple" AAs, according to the fact that AAs are coded by codons that have only pyrimidine in third position ("Y" i.e. "Py" class), and second class with AAs coded by codons with only purine in third position ("R" i.e. "Pu" class). We have the class of "complex" AAs in the case when they are coded by codons that have either pyrimidine, or purine in the third position ("Y or R" i.e. "Py or Pu" class). The said "simplicity" and "complexity" are clearer if we analyze the system NNY & NNR also from the aspect of the rule of the *first-third-letter codon position* (The letter positions I-III: 1. Py-Py, 2. Py-Pu, 3. Pu-Py, 4. PuPu; *see* Tables 3.1 and 3.2).

Now, if we consider the 18 non-sulfur amino acids, we will see that all AAs (all but one, arginine) with "Py or Pu" in third codon position are located on the right side within the system of "5 x 4" (Table 3 in relation to column "Complex combination" in Table 3.2). All "Pu" AAs are located on the left lower side (all except lysine), while those from "Py" class are located on the left upper side (all except histidine). However, like all three base AAs (R, K, H) that are dislocated from "their" groups, two sulfur AAs are dislocated on the same way - they are located on the right side though they should be located on the left, and the "upper" M should go down and "lower" C - up.

After the rule of first-third-letter codon position (Tables 3.1 & 3.2), the right side in Table 3 remains the same (the "complex" AAs), with also dislocated arginine, and with two sulfur AAs (the "simple" AAs) as exceptions. However, all "simple" AAs (all except two sulfur AAs) are on the left side. As we can see, in this new classification, according to the rule of first-third-letter codon position, lysine and histidine are not exceptions any more.

3.2. The Damjanović's spiral model

After all previously mentioned insights, there is a sense for viewing the Solution 4 once more. The Damjanović's spiral model, in relation to Siemion-Siemion's rule, is given in form of "a cross" with two intersecting lines, and with a connection "head to tile" (F-F). By this, the horizontal (shorter) leg consists of AAs of Pu type (K, Q, E, W, M), while the vertical (longer) leg contains two classes: up there are AAs of Py type (italic: F, N, H, D, Y), and down there are AAs of "Py or Pu" type (from T to V-F), with an exeption of cysteine which is of

Py type.

Now there is a question: how is spiral model generated? Approaching Rumer's rule, lead by his own version of "dual logic" (and taking the values of codon bases as: A-0, C-1, G-2, U-3) Damjanović discovered within the codons a "hidden" quaternary numbering system that determines the ordinal number of AAs by reading the letter-doublets ("di-grams") within the codons, from right to left side. [Rumer, 1966, p. 1393: "Considering the group of codons, that relates to one and the same amino acid, shows that within every codon (z | yx) (it should be read from right to left side) it is expedient to separate two-letter 'root' | yx) of the 'end' (z |. So, every amino acid, in a general case, has a corresponding and specific root, and degeneration of the code appears as consequence of exchanging of the endings."]. This way, Damjanović presents a logic from which a series of AAs 0-19 follows, with the interruption of ordinal number 3 (Damjanović, 1998), and also the logic from which he derives the series of AAs 0-21, with interruption of ordinal numbers 3 and 8 (Damjanović, 2003: personal communication).

[The order of AAs through the series 0-19 is as follows: 00 K, 01Q; 02 E; 03 STOP; 04 T; 05 P; 06 A; 07 S; 08 S, R; 09 R; 10 G; 11 C, W, STOP; 12 I, M; 13 L; 14 V, 15 F, L; 16 N; 17 H; 18 D; 19 Y; the series of AAs form zeroth amino acids (0) to the last (21): 00 K, 01Q; 02 E; 03 STOP; 04 T; 05 P; 06 A; 07 S; 08 - ; 09 R; 10 G; 11 C; 12 I; 13 L; 14 V, 15 F; 16 N; 17 H; 18 D; 19 Y; 20 W, 21 M]. In the first case a "two-meaning" logical pattern is presented while in the second case a "one-meaning" logical pattern. For example, the serine is located on two locations (7 and 8) in the "two-meaning" logical pattern and only on one location (7) in the "one-meaning" logical pattern. Accordingly, for arginine, in the second case, we must assume that it is located only in the position 9, since otherwise it would be "mixed" with serine on the position 8. On the other hand, since position 11 must be occupied, with priority, by cysteine, than tryptophan must be moved for one cycle, according to the module 9, and it should appear on the position 20 (11+ 9 = 20). Similar case happens with methionine, which, in relation to isoleucine, moves for one modular cycle further, on the position 21 (12+9 = 21).

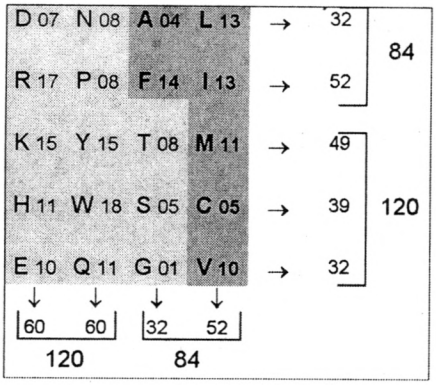
The reading itself (according to Damjanović, 1998) starts with "zero" column where are the codons with middle base "A". Accordingly,

we read zeroth "di-gram" AA as 00 within the codon AAA that is coding for lysine. Subsequently, the neighboring codon AAC, in reverse case as CAA (that is as 100_4), which is the number 16_{10}, as ordinal number for asparagine, etc. In such a manner the 18 AAs are read from the codons. The first to last, the tryptophan, is read from the anticodon (anticodon ACC, read from right to left as CCA) with ordinal number 110_4, that is 20_{10}. (Damjanović, 1998, p. 6: "the cycle of di-grams is presented, and the spiral of codons ... with the 'inverse' appearance of number 20"). Despite the last AA, methionine, Damjanović is read in a specific way from the "ribosomal code", with ordinal number 111_4, that is 21_{10}, it is interesting that ordinal number of methionine can be read also from the "syn-codon", from CCC (which is our hypothesis for further researches). [*Remark 2.* Figure 4 shows that codons with the largest diversity (all three bases are different) together with codons of the smallest diversity (all three bases are the same) contain a balanced number of nucleons. From that it is a reason for introduction of the term "syn-codon" on the following way: UUU is a syn-codon for all six permutations of codon CAG, with the designation UUU/CAG. The same goes for the remaining three syn-codons: CCC/UAG (where one of the permutations is methionine-codon AUG), than AAA/UCG and, finally, GGG/UCA].

3.3. The Davidov's rule

Table 4 contains simultaneously several relations: according to the polarity, according to the first-second-letter codon position, i.e. to the Davidov's rule, and according to the number of atoms. The relation according to the polarity, measured by hydropathy (Kyte & Doolittle, 1982; Doolittle, 1985) and polar requirement Rm (Woese et al., 1966; Rumer & Konopel'chenko, 1975) is presented through columns (quintets), during which Glycine and Proline appear to be ambivalent. They are polar from the aspect of hydropathy, and non-polar from the aspect of polar requirement. Proline ambivalence conditions appearance of a specific "mobile loop" in the system of "5 x 4". Namely, when proline exchanges place with phenylalanine, then appears a sever separation into non-polar (right, bold designated) and polar (left) AAs. Along with that quintets become the bearer of even more subtle distinctions. This way, the first quintet (the first column in Table 4) construct five charged polar AAs (D and E as negatively charged; K,

R and H as positively charged); the second - five polar non-charged; the fourth - five non-polar AAs. If we take in consideration the third quintet (the third column in Table 4), its first "half" is made by non-polar (**G**, A, F) and the second - by polar AAs (**G**, S, T). Besides this, glycine in the first case "plays" as non-polar AA, measured by the polar requirement and in the second case as polar AA, measured by hydropathy.



Tab. 4. *The polar and non-polar amino acids*
The splitting into polar (light ton) and non-polar (bold, dark ton) AAs. The polarity as hydropathy, i.e. as hydropathy index. (After Table 2 in Kyte & Doolittle, 1982, p. 110). The number of atoms within AAs molecules (side chains) in first two rows as well as in second two columns is identical (84 atoms). The same is valid for second three rows in relation to first two columns (120 atoms).

Through the same line by which physical-chemical classification of AAs is being realized on non-polar (right) and polar (left) another strictly chemical classification is being realized also in two categories, according to the Davidov's rule: on the left side there are AAs which in their side chains, apart of carbon, contain oxygen and/or nitrogen, and on the right side there are AAs that contain in their side chains only carbon, or beside carbon also a sulfur. According to their structure proline and glycine (from the aspect of their specific nature) remain out of the classification. [Davydov, 1998, p. 679: "There are two main rules ... (1) all O- or N- ended amino acids ... possess the A-containing

codon doublets; (2) all solely C- ended (excluding Ala) and S- ended amino acids ... have the U-containing codon doublets"... "Pro and Gly are discarded"].

Within the system of "5 x 4" amino-acid pairs are also organized from the aspect of essentiality and non-essentiality. In two "acid" quartets only one pair is essential (**V-L**), and three pairs are non-essential AAs (D-E, N-Q, G-A)[4]. Within three "base" quartets every pair is consisted of one essential and one nonessential or semi-essential AAs (essential- nonessential: **I**-P, **T**-S; essential - semi-essential: **K**-R, **F**-Y, **W**-H, **M**-C).

4. CODON ASSIGNMENT RULES

In the Tables 5-6 it is shown that the assignment of nucleotides and codons to the quartets of amino acid molecules is being realized in accordance with the principle of "the symmetry in the simplest case" (Marcus, 1989). In coding of AAs, the 9 codons (corresponding to modular zero, according to the module 9 in decimal numbering system) participate in the new first quartet (first because K is zeroth AA in Damjanović's model). After this, the number of codons that are coding for the amino acids in the four remaining quartets increases in such

					U	C	A	G				
D	N	A	L	→	12	12	9	9	→	42		
R	F	P	I	→	11	16	9	9	→	45		
K	Y	T	M	→	5	6	13	3	→	**27**	**78**	**78**
H	W	S	C	→	11	10	5	7	→	33		
E	Q	G	V	→	6	4	8	18	→	36		
					45	**48**	**44**	**46**				

Tab. 5. *The distribution of nucleotides number*
Explanation in the text.

[4]This wealth of non-essentiality explains significant role of "acid" quartets, as in biosynthesis of protein AAs and proteins, the same way it is the case in metabolism in general. More about relation of essentiality and non-essentiality of AAs see in Rakočević and Jokić (1996).

a manner that consecutive increments are in the relation 2:1 and 1:2. ("The symmetry in the simplest case"!). Furthermore, the sum of the number of codons and nucleotides, within two inner and two outer quartets, in relation to central quartet, is equal.

					n	c	
K	Y	T	M	→	27	9	
							2
H	W	S	C	→	33	11	
							1
E	Q	G	V	→	36	12	
							2
D	N	A	L	→	42	14	
							1
R	F	P	I	→	45	15	

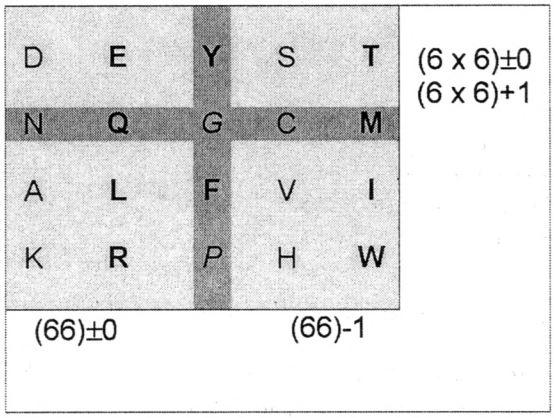
Tab. 6. *The distribution of codon number*
Explanation in the text.

5. THE SYSTEM WITHIN THE SYSTEM

Bearing in mind the chemical composition, from the system of "5 x 4" presented in Table 4, it is possible, through a rearrangement, given a new "4 x 5" system, presented in Table 7. Then, as we can see, in the first row (going from down to up) there are AAs with Carbon and Nitrogen within side chain; in second row AAs with only Carbon, and in the last row AAs with Carbon and Oxygen within side chain. In the first to last row there is a "combination" (Glycine with only Hydrogen; Cysteine and Methionine with Carbon and Sulfur; Asparagine and Glutamine with Carbon, Oxygen and Nitrogen).

As a noteworthy fact in this rearrangement would be an appearance of a specific "cross" where there are AAs which are the "exceptions" from some characteristic aspects: horizontally there are five mentioned combining AAs; vertically: tyrosine as aromatic within aliphatic AAs; Glycine without carbon in side chain; Phenylalanine as aromatic within aliphatic AAs; and, finally, Proline as cyclic aliphatic amino acid. An atom number balance also follows the amino acid arrangement in this new system. Consequently, within horizontal leg of the cross there are (6 x 6)±0 of atoms, and within vertical leg, without glycine, there are

(6 x 6)+1 of atoms within AAs side chains. Without cross, (66)±0 on the left and (66)-1 on the right. At the same time the last quantum (66)-1 is valid in some other AAs splittings. For example, for four AAs in second column plus phenylalanine (E-10+Q-11+L-13+R-17+**F-14** = 65); for four AAs in fifth column plus tyrosine (T-08+M-11+I-13+W-18+**Y-15** = 65); and, finally, for AAs within two columns [(D-07+N-08+A-04+K-15)+(S-05+C-05+V-10+H-11) = 65)]. (Note: If phenylalanine and tyrosine interchange its participation then we would have new solutions: 65±1 and 66±0).



Tab. 7. *A new "4 x 5" AAs system: the splitting into two classes - inner and outer*

This system pursues the system in Table 4. First row (down): AAs with Carbon and Nitrogen within side chain. Second row: AAs with only Carbon within side chain. Last row (up): AAs with Carbon and Oxygen within side chain. First to last row: A combination (Glycine only with Hydrogen; Cysteine and Methionine with Carbon and Sulfur; Asparagine and Glutamine with Carbon, Oxygen and Nitrogen). Within the cross there are only the exceptions: horizontally five the mentioned combining AAs; vertically: Y as aromatic within aliphatic AAs; G without hydrogen; F as aromatic within aliphatic AAs; and, finally, P as cyclic aliphatic amino acid. Within horizontal leg of the cross there are (6 x 6)±0 of atoms, and within vertical leg (without glycine), there are (6 x 6)+1. Without cross: on the left there are (66)±0 and (66)-1 on the right.

6. CONCLUSION

The regularities, presented through previous five Sections provide evidence to support the hypothesis, given in the title of this paper, that in reality it exists a specific harmonic structure of genetic code, shown in Tables 1-7, with coherent functional distinctions, within itself, and with balanced proportionality of atom and nucleon number, as well as of molecules mass of AAs at the same time. On the other hand, the existence of such a harmonic structure with unity of a determination with physical-chemical characteristics and atom and nucleon number at the same time appealed to Aristotle and to his idea of unity of form and essence.

REFERENCES

Damjanović, Z. (1998) Logic core of genetic code, *Glasnik of the Section of Natural Sciences of Montenegrin Academy of Sciences and art (CANU)*, **12**, 5-8.

Davydov, O. V. (1998) Amino acid contribution to the genetic code structure: end-atom chemical rules of doublet composition. *J Theor Biol*, **193**, 679-690.

Dlyasin, G. G. (1998) *Azbuka Germesa Trismegista, ili molekulyarnaya tainopis' myshleniya* , Belye Al'vy, Moskva.

Doollittle, R. F. (1985) Proteins. *Scientific American.* **253**, 74-85.

Konopel'chenko, B. G., Rumer, Yu. B. (1975) Klassifikaciya kodonov v geneticheskom kode, *Doklady Akad. Nauk. SSSR*, **223**, 471-474.

Koruga, Dj., Krstić, J. (1990) Semiconductor and crystal symmetry assessment of microtubule proteins as molecular machines. *J. Mol. Elec.* 6, 167-173.

Koruga, Dj., Tomić, A., Ratkaj, Ž., (2002) Istraživanje biomolekularnih informacionih procesa, Projekt BIOPTRON 3K, ZEPTER INTERNATIONAL, Beograd.

Kyte, J., Doollittle, R. F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105-132.

Marcus, S. (1989) Symmetry in the Simplest Case: the Real Line. *Computers Math. Applic.* **17**, 103-115.

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, *Glasnik of the Section of Natural Sciences of Montenegrin Academy of Sciences and art (CANU)*, **13**, 273-294.

Rakočević, M. M. (2002) Genetic Code: Alphanumerical Regularities Within a Codon Determined Amino Acid Order, *Glasnik of the Section of Natural Sciences of Montenegrin Academy of Sciences and art (CANU)*, **14**, 183-197.

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 sistematizacii kodonov v geneticheskom 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. (1996) A new Manifestation of the arithmetical regularities in Genetic Code, *The 8^{th} ISSOL Meeting and the 11^{th} International Conference on the Origin of Life*, Orleans, France, July 7-12.

Siemion, I. Z., Siemion, P. J. (1994) The informational context of the third base in amino acid codons. *Biosystems*, **33**, 139-148.

Swanson, R. (1984) A Unifying Concept for the Amino Acid Code. *Bull. Math. Biol.* **46**,187-207.

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.

APPENDIX

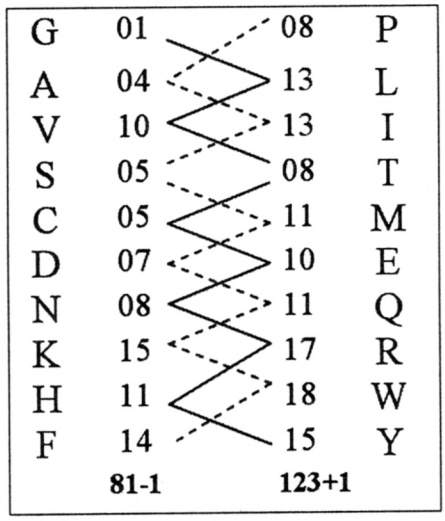


Figure 5. This system of 10 AAs pairs follows from the source system of 12 AAs pairs (Survey 1 in Rakočević & Jokić, 1996, p. 345). On the first (full) zigzag line there are 102-4 whereas on the second (dotted) line 102+4 of atoms. Arithmetic mean for both is 102±4. Separately, within first column (left) at odd/even positions: 35/45 and within second column 61/63 atoms (in source system with 12 AAs pairs 49/49 and 70/68, respectively, what means 69+1/69-1). On the other side, within AAs (side chains) in left column there are 81-1, whereas in the right column 123+1 of atoms. (Cf. result 81/123 in Figures 1; the result 81+1/123-1 in Figure 6 and result 81±0/123±0 in Table 3.6).

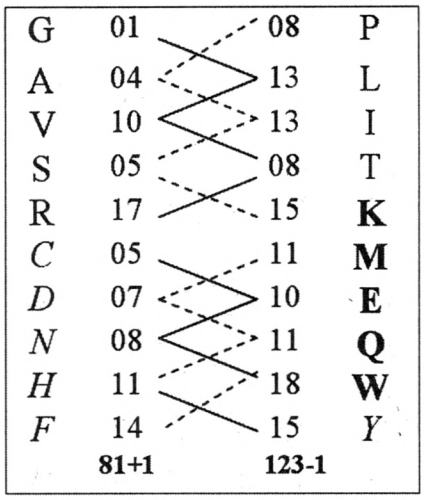


Figure 6. This system of 10 AAs pairs follows from the system presented in Figure 5 and from a splitting into three types of AAs, presented in Tble 3.2. On the first (full) zigzag line there are 102+2 whereas on the second (dotted) line 102-2 of atoms. Arithmetic mean for both is 102±2, in correspondence with 102±4 in system presented in Fig. 5. (*Note:* 2 & 4 as two neighbor numbers in series of even natural numbers; the connection between 2 & 4 exists through three operations: $2+2 = 4$; $2 \times 2 = 4$ and $2^2 = 4$. Furthermore, $2^4 = 4^2$.). On the other side, within AAs (side chains) in left column there are 81+1, whereas in the right column 123- 1 of atoms. (Cf. result 81±0/123±0 in Figures 1, the result 81-1/123+1 in Figure 5 and result 81±0/123±0 in Table 3.6). Within AAs of Py type (italic) there are 60 atoms, of Pu type (bold) 66-1 = 65 and of "Py or Pu" type (common) 37+42 = 78+1 atoms. (Cf. the explanation about the "quantums" 60, 66 and 78 in legend to Table 4.1). Notice the next relations: 60-37 = 65-42 and 42-37 = 65-60. Notice also that the AAs class "Py or Pu" type is splitting into two subclasses: on the left, the AAs (G, A, V, S, R) encoded by codons which possess G in first or second position (one base - G; two positions - I & II); on the right the AAs (P, L, I, T) with codons which possess U or C in second position (two bases - U & C; one position - II) (cf. Table 3.2, down).

II - III	PyPy	PyPu	PuPy	PuPu		
U	38	36	40	18	132	
C	42	42	56	56	196	**(330 – 2)**
A	42	40	26	64	172	
G	28	28	16	22	94	**(330 – 066)+2**
	150	146	138	160		
	[(330–033) – 1]		**[(330–033)+1]**			

Tab. 3.3. *The atom number balances within Table of genetic code after second-third-letter codon position rule (the number of atoms within the rows)*
All is analog with the legend in Table 3.1

[U(Py-Py)]: $F_2(28)+S_2(10)=38$	[U(Py-Pu)]: $L_2(26)+S_2(10)=36$
[C(Py-Py)]: $L_2(26)+P_2(16)=42$	[C(Py-Pu)]: $L_2(26)+P_2(16)=42$
[A(Py-Py)]: $I_2(26)+T_2(16)=42$	[A(Py-Pu)]: $IM(24)+T_2(16)=40$
[G(Py-Py)]: $V_2(20)+A_2(08)=28$	[G(Py-Pu)]: $V_2(20)+A_2(08)=28$
[U(Pu-Py)]: $Y_2(30)+C_2(10)=40$	[U(Pu-Pu)]: $*(00)+*W(18)=18$
[C(Pu-Py)]: $H_2(22)+R_2(34)=56$	[C(Pu-Pu)]: $Q_2(22)+R_2(34)=56$
[A(Pu-Py)]: $N_2(16)+S_2(10)=26$	[A(Pu-Pu)]: $K_2(30)+R_2(34)=64$
[G(Pu-Py)]: $D_2(14)+G_2(02)=16$	[G(Pu-Pu)]: $E_2(20)+G_2(02)=22$

Simple combinations	*Complex combinations*
[(Py-Py)]: F, (60)	[(Py-Py)] or [(Py-Pu)]: L,P,I,T,V
[(Pu-Py)]: Y,H,C,N,D	[(Pu-Py)] or [(Pu-Pu)]: G,R (78+1)
[(Py-Pu)]:**M** (66-1)	[(Py-Py)] or [(Py-Pu)] or [(Pu-Py)]: S
[(Pu-Pu)]:**Q,W,K,E**	[(Py-Py)] or [(Py-Pu)] or [(Pu-Pu)]: –

Tab. 3.4. *The distribution of AAs in correspondence to the second-third-letter codon rule (the number of atoms within the rows)*
All is analog with the legend in Table 3.2

I - II	U	C	A	G		
Py - Py	106	52	-	-	158	**(330 – 2)**
Py - Pu	-	-	74	96	170	
Pu - Py	90	48	-	-	138	**(330 – 066)+2**
Pu - Pu	-	-	80	48	128	
	196	100	154	144		
	[(330–033) – 1]		**[(330–033)+1]**			

Tab. 3.5. *The atom number balances within Table of genetic code after first-second-letter codon position rule (the number of atoms within the columns)*
All is analog with the legend in Table 3.1

[(Py-Py)U]:$F_2(28)+L_6(78)=106$	[(Py-Pu)U]:————
[(Py-Py)C]:$S_4(20)+P_4(32)=52$	[(Py-Pu)C]:————
[(Py-Py)A]:————	[(Py-Pu)A]:$Y_2(30)+H_2(22)+Q_2(22)=74$
[(Py-Py)G]:————	[(Py-Pu)G]:$C_2(10)+W(18)+R_4(68)=96$
[(Pu-Py)U]:$I_3(39)+M(11)+V_4(40)=90$	[(Pu-Pu)U]:————
[(Pu-Py)C]:$T_4(32)+A_4(16)=48$	[(Pu-Pu)C]:————
[(Pu-Py)A]:————	[(Pu-Pu)A]:$N_2(16)+K_2(30)+D_2(14)+E_2(10)=80$
[(Pu-Py)G]:————	[(Pu-Pu)G]:$S_2(10)+R_2(34)+G_4(04)=48$

Simple combinations	*Complex combinations*
[(Py-Py)]:F,L,P **(81)**	[(Py-Py)] or [(Py-Pu)]:–
[(Pu-Py)]:I,M,V,T,A	[(Pu-Py)] or [(Pu-Pu)]:–
[(Py-Pu)]:Y,H,Q,C,W (0101)	[(Py-Py)] or ~~[(Py-Pu)]~~ ~~or~~ ~~[(Pu-Py)]~~ or [(Pu-Pu)]:S (22)
[(Pu-Pu)]:N,**K**,D,E,G	~~[(Py-Py)]~~ or [(Py-Pu)] or [(Pu-Pu)]:R **(0123)**

Tab. 3.6. *The distribution of AAs in correspondence to the first-second-letter codon position rule (the number of atoms within the columns)*
All is analog with the legend in Table 3.2

I - II	PyPy	PyPu	PuPy	PuPu	
U	74	58	-	-	132
C	84	112	-	-	196
A	-	-	82	90	172
G	-	-	56	38	94
	158	170	138	128	
	(330–2)		**[(330–066)+2]**		

(330 − 2)

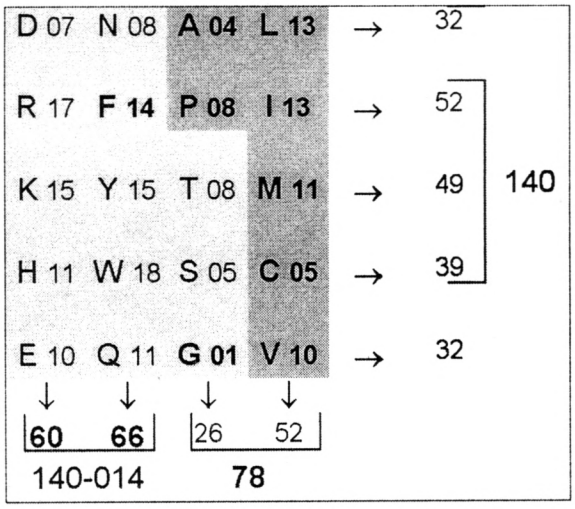
(330 − 066)+2

Tab. 3.7. *The atom number balances within Table of genetic code after first-second-letter codon position rule (the number of atoms within the rows)*
All is analog with the legend in Table 3.1

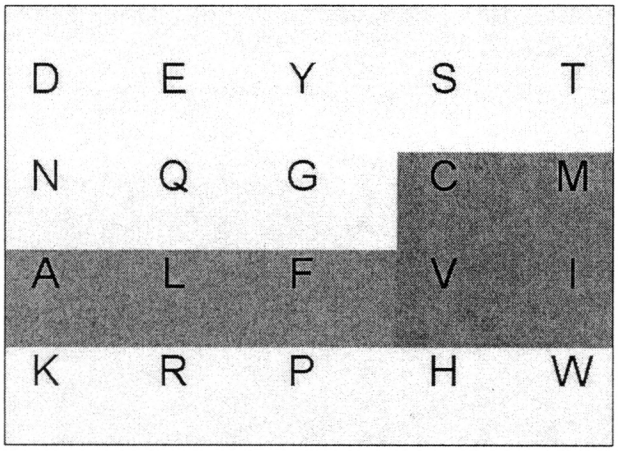
[U(Py-Py)]:$F_2(28)+L_2(26)+S_4(20)=74$	[U(Py-Pu)]:$Y_2(30)+C_2(10)+W(18)=58$
[C(Py-Py)]:$L_4(52)+P_4(32)=84$	[C(Py-Pu)]:$H_2(22)+Q_2(22)+R_4(68)=112$
[A(Py-Py)]:———	[A(Py-Pu)]:———
[G(Py-Py)]:———	[G(Py-Pu)]:———
[U(Pu-Py)]:———	[U(Pu-Pu)]:———
[C(Pu-Py)]:———	[C(Pu-Pu)]:———
[A(Pu-Py)]:$I_3(39)+M(11)+T_4(32)=82$	[A(Pu-Pu)]:$N_2(16)+K_2(30)+S_2(10)+R_2(34)=90$
[G(Pu-Py)]:$V_4(32)+A_4(16)=56$	[G(Pu-Pu)]:$D_2(14)+E_2(20)+G_4(04)=38$

Simple combinations	*Complex combinations*
[(Py-Py)]:F,L,P **(81)**	[(Py-Py)] or [(Py-Pu)]:–
[(Pu-Py)]:I,M,V,T,A	[(Pu-Py)] or [(Pu-Pu)]:–
[(Py-Pu)]:Y,H,Q,C,W (0101)	[(Py-Py)] or ~~[(Py-Pu)]~~ or ~~[(Pu-Py)]~~ or [(Pu-Pu)]:S (22)
[(Pu-Pu)]:N,**K**,D,E,G	~~[(Py-Py)]~~ or [(Py-Pu)] or [(Pu-Pu)]:R **(0123)**

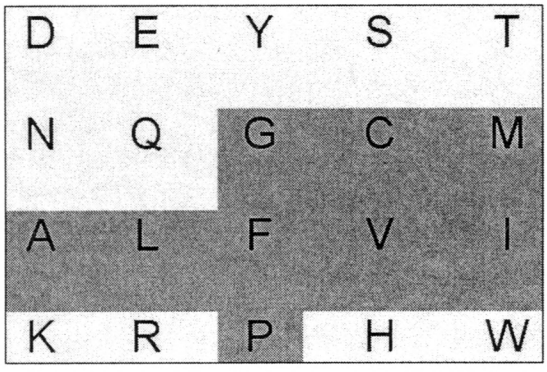
Tab. 3.8. *The distribution of AAs in correspondence to the first-second-letter codon position rule (the number of atoms within the rows)* All is analog with the legend in Table 3.2



Tab. 4.1. *The changes in atom number balances*
This Table pursues Table 4 with one interchange within "mobile F-P loop". Within five polar charged AAs (side chains), their five neighboring AAs, and within last ten AAs there are 60, 66 and 78 atoms, respectively; all this, the same as within seven "golden" AAs, seven (their) "complements", and six "non-complements". (Cf. Survey 2.1 in Rakočević, 1998, p. 289). Notice that the differences are 1 x 6, 2 x 6 and 3 x 6 (the validity of principles of continuity and minimum change).



Tab. 7.1. *The splitting into polar and non-polar AAs after hydropathy*
The splitting into polar (light tone) and non-polar (dark tone) AAs. (The values for hydropathy index from Table 2 in Kyte & Doolittle, 1982, p. 110).



Tab. 7.2. *The splitting into polar and non-polar AAs after polar requirement*
The splitting into polar (light tone) and non-polar (dark tone) AAs. [The values for polar requirement from Table 2 in: Woese et al., 1966, p. 731; and from Solutions (4) & (5) in: Konopel'chenko & Rumer, 1975, p. 473].

	CG	UA						UA	CG	
421 (1443)	**146 (524)**	**275 (919)**	DGA	EGA	YUA	SUC AG	TAC	**182 (652)**	**254 (897)**	436 **(1550-1)**
365 (1328)	*149 (564)*	*216 (764)*	NAA	QCA	GGG	CUG	MAU	*253 (899)*	*184 (675)*	437 (1573+1)
	295 (1088)	491 (1683)	AGC	LUU CU	FUU	VGU	IAU	435 (1550+1)	438 (1573-1)	
C & combin.			KAA	RCG AG	PCC	HCA	WUG	(N & O)		

Tab. 7.3. *The atom and nucleon number balances within base doublets of codons coding for AAs*

In middle position is the system of "4 x 5" of AAs as in Table 7, but here with their base doublets of corresponding codons. On the right - the data for nitrogen (first row, going from up to down, bold: 182 & 254 of atoms and 652 & 897 of nucleons within codon doublets of AAs: K,R,P,H,W), and data for oxygen AAs (second row, italic: 253 & 184 of atoms and 899 & 675 of nucleons within codon doublets of AAs: D,E,Y,S,T). On the left - the data for carbon AAs (first row, going from up to down, bold: 146 & 275 of atoms and 524 & 919 of nucleons within codon doublets of AAs: A,L,F,V,I), and for "combining" AAs (second row, italic: 149 & 216 of atoms and 564 & 764 of nucleons within codon doublets of AAs: N,Q,G,C,M). As it is self- evident there are strict regularities in atom number and nucleon number as a connection between nitrogen and oxygen AAs (436 versus 435, and 437 versus 438 of atoms; similarly is for nucleon number). (*Note:* The number of atoms within nucleotides: UMP-34, CMP-35, AMP-37, GMP-38 and the number of nucleons: U-112, C-111, A-135, G-151).