

# A “Quantum-Like” Approach to the Genetic Code

Tidjani Négadi

## Abstract

A “quantum-like” analogy, introduced in 1998 by Rakočević in his study of the biosynthetic amino acid precursors, with the prime “quantum O37” in the background, is also considered in this paper as a guiding line of thought. We use and extend his results to derive (i) the multiplet structure of the standard genetic code, in a completely new and original way and (ii) several physicochemical quantities, as the number of nucleons in the 61 coded amino acids, the number of nucleons in 20, 23 and 61 amino acid biosynthetic precursors, the number of nucleons in 23 amino acids signals (AASs), the number of atoms in the 64 DNA-codons and in the 64 RNA-codons, and several other interesting quantities. In this way, we show that the fundamental set of seven fundamental amino acid precursors grouped into six amino acid families appears more informative than the larger set of twenty canonical amino acids. In a second part, we find once more a conspicuous numeric “manifestation” of serine, its protonated/neutral octamers and several other serine clusters; these clusters, in particular the serine octamer are thought to have been key players in the origin of homochirality. Also, we find, encoded, water (H<sub>2</sub>O) and, interestingly, the exact number of its characteristic water-water Hydrogen-Bond Contact/Non-Contact configurations (or “degeneracies”) according to the body-centered cubic (bcc) lattice gas model for water.

**Key Words:** amino acid precursors; genetic code multiplets; serine clusters, water

NeuroQuantology 2011; 4: 785-798

## 1. Introduction

For several years, in our study of the genetic code, we have been focused mainly on the 20 (canonical) amino acids coded by it. Now, the study can be fruitfully extended to include the smaller and maybe more fundamental set of amino acid biosynthetic precursors (Négadi, 2011). It is a well known fact that the whole process of amino acid biosynthesis in living organisms relies on a small set of only seven amino acid precursors and the 20 amino acids are grouped into six families according to their metabolic precursors (see Table 2 and, for example,

Berg *et al.*, 2011). The precursors of the amino acids are: 1) Ribose-5-phosphate, 2) 3-Phosphoglycerate, 3) Pyruvate, 4) 2-Oxoglutarate (or  $\alpha$ -Ketoglutarate), 5) Phosphoenolpyruvate+Eritrose-4-phosphate and 6) Oxaloacetate. It is clear that the real biosynthesis processes are incredibly more complex than this simple “reduction”, as there are myriads of other molecular intermediate (enzymes, metabolites, etc.) and processes, but even starting with such a “drastic” reduction to only six precursors, will be shown to be highly interesting and informative. Thirteen years ago, Rakočević (Rakočević, 1998) derived interesting whole-number relations between amino acids and their biosynthetic precursors, confirming the existence of *regularities* based on the “quantum” 037 (shCherbak, 2008). Specifically, by distinguishing the three components (molecular parts) of the

Corresponding author: Tidjani Négadi, Dr.

Address: Département de Physique, Faculté des Sciences, Université d’Oran, 31100, Oran, Algérie

Phone: +213 0795859481

e-mail: tnegadi@gmail.com

Received Jun 30, 2011. Revised Sept 20, 2011.

Accepted Dec 7, 2011.

precursor molecules: 1) the "head", 2) the "body" and 3) the "tail", Rakočević derived the same kind of regularities valid for the amino acids themselves where, also, a distinction has frequently been made between the "side-chains" and the "block". He then computed the *atom number* of amino acid biosynthetic precursors (that is atom number in the *precursors molecules*) for the 23 amino acids signals (AASs), i.e., 8 four-codon amino acids (5 quartets and 3 quartet-parts of the three sextets, p-fcaa in the Table 1) and 15 non-four-codon amino acids (the 9 doublets, the 3 doublet-parts of the 3 sextets, the triplet and the two singlets, p-nfcaa). There are 111 atoms in the 8 four-codon amino acid precursors ("head"+"body"+"tail") and  $(68+86)+138=224+68=222+2+68=292$  atoms in the non-four-codons amino acid precursors ( $\{{"head"}+{"body"}\}+{"tail"}$ )<sup>2</sup>, for a total of 403 atoms (Table 1).

**Table 1.** The distribution of atom number of amino acid precursors between "head", "body" and "tail" (from Rakočević, 1998; with correction made) Symbols: p-nfcaa: precursors of non four codon amino acids; p-fcaa: precursors of four codon amino acids.

|        | p-nfcaa      | p-fcaa    | S  | S'           |
|--------|--------------|-----------|----|--------------|
| "head" | 68           | 32        | 36 |              |
|        | ↑"2"         |           |    | 30           |
| "body" | <b>86</b>    | <b>20</b> | 66 |              |
|        |              |           |    | 13           |
| "tail" | <b>138</b>   | 59        | 79 |              |
|        | <b>222+2</b> | 111       |    | <b>222+2</b> |

In the above Table, S denotes the sequence of *Differences*, S' the sequence of *Difference of Differences* (and for later use we also use S" the sequence of *Difference of Differences of Differences*). The discovery of a "providential" error in the paper by Rakočević [Table 1 in: (Rakočević, 1998)], mentioned above, led us to reconsider the problem and to show that, indeed, the pattern 222 could emerge, *two times*, at the sole condition to apply a shCherbak-like procedure of "borrowing" 2 atoms, for example, from the "body" to the benefice of the "head" in the set p-nfcaa (the arrow in the Table 1). In this way, it is not difficult to show that the three patterns 222, 111 and 222, occupying respectively the last row in the Table 1, emerge. The second pattern 222

<sup>2</sup> There is a number in this reference in need of fine-tuning (see Table 1 of Rakočević, 1998: the "body"-part of atom-number of amino acid precursors is checked as equal to 86 (not 84).

above has been obtained using what Rakočević calls the sum of "*Maximum Possible Changes*" (MPC), i.e., *Differences plus Differences of Differences*. We have also found another interesting case where the pattern 222 appears again and *two times*: the carbon atom-number (see details in Section 3). The number of *carbon atoms* in the 61 amino acid precursors is, from Table 2, 249 and the *Difference* between this last one and the number of carbon atoms in the six precursors, 27, is

$$\#C\text{-atoms in 61 aa-precursors})-(\#C\text{-atoms in 6 aa-precursors})=222$$

In the case of the nucleon-number of amino acid biosynthetic precursors, he showed that the number 2211, playing here the role of 222 in the atom-number case, also appears two times, first as the number of nucleons in the 23 amino acids signals ("body", only, for four-codon amino acid precursors and "head"+"tail" for non-four codons amino acid precursors) and also and foremost as the "*Maximum Possible Changes*", much the same situation as in the above case for atom-numbers (see the corresponding Table 3). Note, importantly, that there are 1060 nucleons in the 8 four-codon amino acids precursors and 2892 in the 15 non-four-codon amino acid precursors (total for 23: 1060+2892=3952). We shall return to this latter sum, later.

The existence of these remarkable regularities led Rakočević to use the vocabulary of quantum physics, *as analogy*, to describe these striking regularities, revolving around the two patterns, or "orbitals", 222 and 2211, as he calls them. He said, for example, "*namely, we can understand that one of two "quanta" (whole numbers) from the third position in the first pattern has been "promoted" (2=1+1) and occupies a previously vacant "orbital" in the second pattern.*" (One could, fancily, think of numbers, expressed in the place-value numerical system, as abstract "atoms" where the "place-values", the "orbitals", could be either vacant (zero) or "occupied" (with one of nine digits 1-9, in the decimal system) and the arithmetic operations between numbers play the role of "interactions".) In this paper, we shall willingly adopt this "quantum-like" analogy inasmuch as in the vector space of

(number) sequences of finite difference calculus the iterative action of the linear difference operator on sequences “resemble” the action of the derivatives in differential calculus. We thus consider the two “states” or “orbitals”: the first, **222**, as “characterizing” *carbon-numbers and/or atom numbers*, as we shall precise below, and the second, **2211**, as “characterizing” *nucleon-numbers*. We shall take these two “states” as a starting point.

The work is organized as follows. In Section 2, we use the above patterns as well as other derived quantities to establish an equation with four integer parameters, our Eq.(3), encoding, at the same time, various nucleon-numbers (61 amino acids, 20, 23 and 61 amino acid precursors) and also, through the parameters values, the multiplet structure of the standard genetic code. In Section 3, we first extend the study to *carbon-atom composition* and derive new difference and difference of difference relations, for later use and, next, establish from the results of Section 2 a remarkable set of three 5-digits numbers describing again, and in a very elegant way, the multiplet structure of the genetic code. The last Section is mainly devoted to the presentation of what we may call the “epiphany” of serine and its “magical-number” clusters, as experienced in the context of this paper. (We recall that the

latter “manifestation” occurred, also and recently, in the context of our *irregular (integer) tetrahedron* classification model of the 20 amino acids (Négadi, 2011a). Finally, some remarks on the possibility of (a fruitful) inclusion of water in the study of the genetic code are made, inasmuch as the “manifestation” of water resembles serine’s one.

## 2. The multiplet structure of the genetic code from the amino acid precursors

Let the *Difference* MPC(*nucleons*)-MPC(*atoms*) defines the new “orbital”

$$\Psi := 2211 - 222 \quad (1)$$

which numeric value is 1989. Each one of the two “states” **2211** and **222** is the *sum of differences and Difference of Differences*. The above “state”,  $\Psi$ , a simple number, is nevertheless interesting as the application of Euler’s “operator”  $\phi$  to  $\Psi$  (see below the definition of the function  $\phi$ ) gives  $\phi(\Psi) = 1152$ , and the latter value is simply related to the *total number of nucleons in the precursors* (just two “quanta” below  $\phi(\Psi)$ , see the Table below):

$$\phi(\Psi) - 2 = 1150 \quad (2)$$

**Table 2.** The detailed atomic composition (carbon C, oxygen O, phosphorus P and hydrogen H) of the amino acids precursors; the respective distribution between “head”, “body” and “tail” is given, in the parenthesis.

| precursor              | C        | O  | H  | P | #atoms      | #nucleons       | Amino acids      |
|------------------------|----------|----|----|---|-------------|-----------------|------------------|
| 1. Ribose-5-phosphate  | 5(1,3,1) | 8  | 11 | 1 | 25(3,12,10) | 230(29,90,111)  | H                |
| 2. 3-Phosphoglycerate  | 3(1,1,1) | 7  | 7  | 1 | 18(4,4,10)  | 186(45,30,111)  | G, S, C          |
| 3. Pyruvate            | 3(1,1,1) | 3  | 4  | 0 | 10(4,2,4)   | 88(45,28,15)    | A, L, V          |
| 4. 2-Oxoglutarate      | 5(1,1,3) | 5  | 6  | 0 | 16(4,2,10)  | 146(45,28,73)   | P, E, Q, R       |
| 5. Phosphoenolpyruvate |          |    |    |   |             |                 |                  |
| +Eritrose-4-phosphate  | 7(2,3,2) | 13 | 14 | 2 | 36(7,16,13) | 368(74,169,125) | W, F, Y          |
| 6. Oxaloacetate        | 4(1,1,2) | 5  | 4  | 0 | 13(4,2,7)   | 132(45,28,59)   | T, M, I, D, N, K |
| Total                  | 27       | 41 | 46 | 4 | 118         | 1150            | 20/23            |

We shall return to the number 1152, in the last Section, where it will be shown that it hides more information. We now construct from  $\Psi$  new “states”, using the “operators”  $\sigma(\cdot)$ ,  $\phi^{(2)}(\cdot)$ ,  $\phi^{(3)}(\cdot)$  and  $\mathbf{a}_0(\cdot)$ , where  $\sigma$  is the sum of divisors function,  $\phi$  is Euler’s function which gives the number of numbers coprime to a given number,

together with its first few iterations  $\phi^{(i)}$  and finally  $\mathbf{a}_0$  is the sum of prime factors arithmetic function. We obtain  $\sigma(\Psi) = 3276$ ,  $\phi^{(2)}(\Psi) = 384$ ,  $\phi^{(3)}(\Psi) = 128$ ,  $\mathbf{a}_0(\Psi) = 36$ .  $\mathbf{v}(\Psi) = \frac{\mathbf{a}_0(\Psi)}{\sqrt{\mathbf{a}_0(\Psi)}} = 6 (=222/37)$ .

The above particular set of six numbers, including  $\phi(\Psi)$ , proves to be full of

interesting properties and in the following only but one of them is examined. First, there is a "large" component,  $\sigma(\Psi)=3276$ , and a series of "small" components constituted by the remaining five numbers with sum 1706 made of two parts the larger  $\phi$ -part (1664) and the smaller  $\mathbf{a}_0$ -part (42). We have that the *mean* of the two parts is equal to 853 and twice their sum gives 3412. This latter number is equal to the number of nucleons in the 61 coded amino acids, in their "physiological" state (i.e., some aas in ionized state and proline's modification meant, see shCherbak, 2008; see also Négadi, 2009, 2011).

As for the former 853, it is also related to our recent work on the amino acid precursors (Négadi, 2011a) where we derive, in particular, and only from elementary functions of it, the total number of atoms in the 61 amino acids RNA-codons 2560 and also the number of amino acids 3412, mentioned above. Other interesting derivations from the number 853 could be found in the above reference, where this number is found another way. We shall make an important remark on this number in the last Section. The second iteration  $\phi^{(2)}(\Psi)$  appears identical with the number of atoms in the 20 amino acids (side-chain and block) and the third,  $\phi^{(3)}(\Psi)$  is simply 27. Let us consider now the following general linear combination of the four "states", constructed above, with  $\alpha$ ,  $\beta$ ,  $\gamma$  and  $\delta$ , a set of *integer* coefficients

$$\bar{\Psi}_{(\alpha,\beta,\gamma,\delta)}(\Psi) = \alpha\sigma(\Psi) + \beta\phi^{(2)}(\Psi) + \gamma\phi^{(3)}(\Psi) + \delta\mathbf{a}_0(\Psi) \quad (3)$$

The following five sets of coefficients, in cases (o) and (i)-(iv), were found interesting, as they reproduce nicely several meaningful numbers with regard to the precursors, the amino acids and the total "degeneracy":

(o)  $\alpha=0, \beta=1, \gamma=0, \delta=0$  (and  $\alpha=0, \beta=0, \gamma=3, \delta=0$ ):

$$\bar{\Psi}_{(0,1,0,0)}(\Psi) = \bar{\Psi}_{(0,0,3,0)}(\Psi) = 384 \quad (4)$$

(i)  $\alpha=1, \beta=0, \gamma=1, \delta=0$ :

$$\bar{\Psi}_{(1,0,1,0)}(\Psi) = 3404 \quad (5)$$

(ii)  $\alpha=1, \beta=0, \gamma=2, \delta=0$ :

$$\bar{\Psi}_{(1,0,2,0)}(\Psi) = 3532 \quad (6)$$

(iii)  $\alpha=1, \beta=0, \gamma=5, \delta=1$ :

$$\bar{\Psi}_{(1,0,5,1)}(\Psi) = 3952 \quad (7)$$

(iv)  $\alpha=1, \beta=0, \gamma=49, \delta=3$ :

$$\bar{\Psi}_{(1,0,49,3)}(\Psi) = 9656 \quad (8)$$

The first case, (o), corresponds to the total number of atoms in the 20 amino acids (side-chain-and-block),  $204+9 \times 20$ , mentioned above and a well-known result. Note, importantly, that 384 appears *two times* which means that we dispose of an identity " $384-384=0$ " which will prove useful in the following applications (see below). The next four cases, (i)-(iv), are concerned, *all*, with *nucleon-numbers* of the precursors *and* those of the amino acids and, as we shall see right now, not only they (the cases (i)-(iv)) give, each, a meaningful number, but we show also that their grouping constitutes a remarkable result: *the sum of all the integer coefficients in Eqs.(5)-(8) is equal to the total number of the amino acid coding-codons and the three stop-codons* in a beautifully appropriate and, a-priori, unexpected manner.

Before examining the last point, let us return briefly to Eqs.(5)-(8). The first case, (i) and Eq.(5), is seen to be equal to the total number of nucleons in 61 amino acids coded, again a well known result (shCherbak, 2003; see also Négadi, 2008, 2009, 2011a). The case (ii) and Eq.(6) is equal to the total number of nucleons in 20 *amino acid precursors*

$$230 + 3 \times 186 + 3 \times 88 + 4 \times 146 + 3 \times 368 + 6 \times 132 = 3532, \text{ see the Table 2).}$$

The case (iii) and Eq.(7) is equal to the total number of nucleons in 23 *Amino Acid Signal precursors*  $230 + (3+1) \times 186 + (3+1) \times 88 + (4+1) \times 146 + 3 \times 368 + 6 \times 132 = 3952$ , see the Table 2). In the latter sum, the position of the second contribution of the three sextets is highlighted in bold. As for the last case, (iv) and (8), a short calculation using the same kind of reasoning<sup>3</sup>

<sup>3</sup>  $(2 \times 230) + (12 \times 186) + (14 \times 88) + (14 \times 146) + (5 \times 368) + (14 \times 132) = 9656$ .

shows that it is equal to the total number of nucleons in the 61 amino acid precursors molecules: 9656.

On the basis of the remarkable existence of the above set of four "orbitals", each one of them describing correctly a meaningful situation with respect to both the sub-atomic composition (nucleon-numbers) and also with respect to the genetic code degeneracy to which we return now. Specifically, by adding all the (integer) coefficients in the four (nucleon-number) cases ("orbitals"), we obtain

$$\sum_{(i)}^{(iv)} \gamma + \sum_{(i)}^{(iv)} \delta = (1 + 2 + 5 + 49) + \quad (9)$$

$$(1 + 3) = 56 + 3 + 1 + 1 = 61$$

a number equal to the total number of codons. Also, we have two possible values for the coefficients of the unique case, (o), which concerns atom-number only:  $\beta=1$  and  $\gamma=3$  and they are related because  $\varphi^{(2)}(\Psi)=3 \times \varphi^{(3)}(\Psi)$ . Before explaining the result in (9) for the 61 codons, we could, first, take  $\gamma=3$  as representing the number of stops (standard genetic code):

$$\#stop\ codons = \gamma = 3 \quad (10)$$

The first sum, in Eq.(9), gives a final value that seem to reproduce nicely the total number of amino acid coding codons 61 as 56 codons for the three even-degeneracy classes and 5 codons, in the form 3+1+1, for the odd-degeneracy classes, that is 3 for the triplet and 1+1 for the two singlets. In this way the total number of codons, from Eqs.(9)-(10), is given by 61+3=64. In a private communication (Rakočević, 2011) Rakočević asked the question "6 or 7" about the number of precursors (Négadi, 2011), depending on whether or not the two (related precursors) Phosphoenolpyruvate and Eritrose-4-phosphate are considered as one unique precursor (Rakočević, 1998) or two separate precursors (Négadi, 2011b, for a possible "6 and/or 7" answer). The latter possibility could be exploited to write the following beautiful and simple relation connecting 6 and 7:  $6^2+(6+7)=7^2=49$  and next inject it in Eq.(9) to obtain, modulo an

elementary rearrangement, using  $36 = a_o(\Psi) = (3+17) + (3+13)$

$$(20 + 18 + 18 + 3 + 2) + 3 \quad (11)$$

$$= 61 + 3 = 64$$

which gives finally the correct number of codons in each one of the five degeneracy classes of the standard genetic code (5 quartets: 20, 9 doublets: 18, 3 sextets: 18, 1 triplet: 3 and 2 singlets: 2) and of course the 3 stop-codons. Note that Eq.(11) could also, and easily, be represented in the form

$$[(20 + 12) + (18 + 6)] + (3 + 2) + 3 \quad (12)$$

$$= 61 + 3 = 64$$

where 20=2+18 is associated (total number of codons) to the five quartets, 12=5+7 is associated to the total number of codons in three quartet-parts of the three sextets and finally 24=18+6 corresponds to the number of codons in the nine doublets (18) and in the three doublet-parts of the three sextets (6). The parentheses separate the quartets from the doublets, the even classes, on the one hand, from the odd classes (3+2), on the other. The set of amino acids corresponding to the part 20+12 is what we called four-codon set and those of the part (18+6)+(3+2) non-four-codon set. To support the two-fold classification of the amino acids into "four-codon" and "non-four-codon" classes, let us compute the sum of  $\Psi$ , from Eq.(1), and the number of nucleons in the six precursors in the Table 2 as well as their number. We find 3145, obtained as 1989+1150+6 or (2211+1156)-222=3367-222, or else 2211+934. If only 9 is introduced (in the form of the identity -9+9=0) in the latter relation, we could reach 2220+925=3145 which is the partition of nucleon numbers into respectively the non-four-codons class and the four-codon class (60x37 and 25x37). The above identity -9+9=0 could be found from  $\varphi(\Psi)/\varphi^{(3)}(\Psi) = csod(1989) = 9$  where csod is the complete sum of digits. We shall return to the above partition in the next Section.

### 3. The multiplet structure of the genetic code, again, but from a number

Before beginning the "body" subject of this Section, let us return briefly to the number of *carbon atoms* in the 61 amino acid precursors, mentioned in the introduction. It is possible to establish for carbon atom-numbers an analogous "difference of difference" relation. Writing  $C_{aa-p}^{(61)}=249$  for the total number of *carbon atoms* in 61 amino acid precursors (see the Introduction),  $C_{AAS-p}^{(23)}=99$  the total for 23 amino acid signals and finally  $C_p^{(6)}=27$  for the total number of carbon atoms in the 6 precursors (see Table 2). Below, we first compute the "differences", in Eqs.(13) and (15), and next the difference of differences, in Eq.(14).

$$C_{aa-p}^{(61)} - C_{aa-p}^{(23)} = 150 \quad (13)$$

$$72 \quad (14)$$

$$C_{aa-p}^{(61)} - C_p^{(6)} = 222 \quad (15)$$

Several interesting conclusions could be made. First, the sum of differences and difference of differences gives  $222+(150+72)=222+222=444$ . Second, the sum of differences, only, gives 372 ( $150+222$ ). It appears that this last number is identical with  $H_{aa-p}^{(61)}=372$ , the total number of *Hydrogen atoms* in 61 amino acid precursors. Some hints made recently by Rakočević in a private communication, led us to go beyond the above obtained result,  $150+222$ , and call the "borrowing" relation  $222+2=224$  (see the introduction) and inject it into the sum of (13) and (15) to get finally  $148+224=372$  which is the right hydrogen-number pattern for odd order and even order precursors, respectively [see Table 2 above and Table 1 in: (Négadi, 2001b)]. Third, working still with hydrogen, we find that the difference

$$H_{aa-p}^{(61)} - H_p^{(6)} = 372 - 46 = 326 \quad (16)$$

a number which appears equal to the number of atoms in the three stops (Rakočević, 2009).

After these remarks, we shall exploit, here, some of the results of the second Section to construct a highly interesting number (in fact three inter-connected

numbers) which describes, remarkably, all aspects of the genetic code degeneracy. Consider the following three term sum

$$\bar{\Psi}_{(1,0,49,3)}(\Psi)+\sigma(\Psi)+6=12938 \quad (17)$$

The first term  $\bar{\Psi}(1,0,49,3)=9656$  (the largest component, see above) is the total number of nucleons in the 61 *amino acid precursors molecules* (Eq.(8)). The second, equal to 3276, is the sum of the divisors of  $\Psi$ ,  $\sigma(\Psi)$ , and the third, 6, is the number of the fundamental precursors. The total sum, the number 12938, is a *very interesting number*. Without adding the number of precursors, 6, we would obtain 12932. This latter and also the mean  $(12932+12938)/2=12935$  are also *both interesting*. For the mean, for example, and we shall begin by examining this case, first, *each digit is associated to the number of amino acids in each one of the five well known multiplets of the (standard) genetic code: 1 for the triplet, 2 for the singlets, 9 for the doublets, 3 for the sextets and finally 5 for the quartets and, of course, a total sum equal to 20*. Let us see how to extract the "degeneracies" in this latter number (12935) and, next, return to the other two interesting cases (12938 and 12932), showing in passing that all *three cases are very closely related and each one of them describes a facet concerning the ways the different multiplets are grouped*. First, highlight the "exponents" in the decimal representation of 12935:

$$\begin{array}{cccccc} & 4 & 3 & 2 & 1 & 0 \\ \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow \\ \mathbf{1} & \mathbf{2} & \mathbf{9} & \mathbf{3} & \mathbf{5} & \end{array} \quad (18)$$

We see at once that for the **9** doublets, there is *perfection*: the number of codons in an individual doublet is just its associated exponent **2** (9 is multiplied by  $10^2$ ) so that the product  $9 \times 2 = 18$  gives exactly the number of codons. The **3** sextets and the **5** quartets, on the other hand, are linked because each one of them (i.e., the digits **3** and **5**) gives the number of degenerated codons in an *individual* multiplet of the other: a sextet has 5 degenerated codons and a quartet has 3 degenerate codons. Accordingly, by associating the exponent 1 to both numbers, gives finally the number of codons itself: for the **5** quartets we get  $5 \times (1+3) = 20$  and for the **3** sextets

$3 \times (1+5) = 18$ . In the three other odd-class amino acids, the exponents are not necessary and we obtain the number of codons in a very elementary way: for the **2** singlets it is the simple product of their number (**2**) with the number for the (unique) triplet (**2** $\times$ **1**), and for the triplet, it is the sum of this latter and the number of doublets (**1** $+$ **2**). Below, we sum-up the above results which reproduce exactly the number of codons in each one of the five multiplets of the standard genetic code:

$$\begin{array}{l}
 9 \text{ doublets: } 9 \times e(9) = 9 \times 2 = 18 \\
 5 \text{ quartets: } 5 \times [e(3)+3] = 5 \times (1+3) = 20 \\
 3 \text{ sextets: } 3 \times [e(3)+5] = 3 \times (1+5) = 18 \\
 2 \text{ singlets } 2 \times 1 = 2 \\
 1 \text{ triplet } 1 + 2 = 3
 \end{array} \quad (19)$$

We have thus shown that the number 12935 describes correctly the five degeneracy classes, with a total of 61 codons. As an important remark, to be further exploited in Section 4, we have that the position of the digit 9 is "strategic" because of its association to the exponent 2 and a change in position by permutation would destroy the validity of the calculation of the degeneracies. Also, we have clearly that the digits 1 and 2, on the one hand, and 3 and 5, on the other, are *permutable* without affecting the result. Next, we consider the other number 12938. Remarkably, this number, shown below also with highlighted exponents, on the other hand, describes nicely and in the detail the two-fold four-codon/non-four-codon partition. As in the preceding case, but here, the sum of the five digits:  $(1+2+9+3)+8=15+8=23$ , just the right numbers of Rumer's division 15 AASs in the non-four-codon set and 8 AASs in the four-codon (Rumer, 1966).

$$\begin{array}{cccc}
 4 & 3 & 2 & 1 & 0 \\
 \hat{1} & \hat{2} & \hat{9} & \hat{3} & \hat{8}
 \end{array} \quad (20)$$

Here, the digit 3 is related to the digit 9, not to the digit 8 as in the preceding case where 5 replaced 8. Nine is still the number of doublets, *as before*, and the digit 3 corresponds now to the 3 doublet-parts of the three sextets ( $S^{II}, L^{II}, R^{II}$ ). The last digit 8 (as  $3+5$ , see below) is associated to the four-codon set (*fc-set*) with 8 quartets (the 5

ordinary quartets and the 3 quartet-parts of the three sextets  $\{S^{IV}, L^{IV}, R^{IV}\}$ ). Finally, for the two singlets (2) and the triplet (1), there is no change with respect to the preceding case. Using the same kind of reasoning as above, we get finally the following pattern which is also the correct multiplet composition of the two sets of four-codon and non-four-codon amino acids with a total of respectively 32 and 29 codons.

$$\begin{array}{l}
 (5+3) \text{ quartets} + (S^{IV} + L^{IV} + R^{IV}): \quad (21) \\
 8 \times [e(3)+3] = 8 \times (1+3) = (5+3) \times (1+3) = 20 + 12 = 32 \\
 9 \text{ doublets: } 9 \times e(9) = 9 \times 2 = 18 \\
 3 \text{ doublet-parts } S^{II} + L^{II} + R^{II}: 3 \times e(9) = 3 \times 2 = 6 \\
 2 \text{ singlets: } 2 \times 1 = 2 \\
 1 \text{ triplet: } 1 + 2 = 3
 \end{array} \quad \left. \vphantom{\begin{array}{l} \\ \\ \\ \\ \\ \end{array}} \right\} \text{nfc-set}$$

Now, we return to the number 12932 which is obtained by dropping the number 6 of precursors. We have

$$\begin{array}{cccc}
 4 & 3 & 2 & 1 & 0 \\
 \hat{1} & \hat{2} & \hat{9} & \hat{3} & \hat{2}
 \end{array} \quad (22)$$

The sum of the digits in this case, written as  $1+2+9+(3+2)=17$ , suggests, and we shall confirm below by computing the "degeneracies", that the 6 AASs  $S^{IV,II} + L^{IV,II} + R^{IV,II}$  are missing and the remaining 17 amino acids are those with no degeneracy at the first-codon position. We recall that serine, leucine and arginine are all three doubly degenerate at the *first-codon position*. It is also striking that the *missing* of the 6 AASs associates with the *exclusion* of the 6 precursors from Eq.(17). Here, the digit 3 works two times, one time added to its neighbor digit 2 to give the number of quartets 5, and the other time, it is added to its own exponent 1 to give 4 ( $=3+1$ ) the number of codons in a single quartet. For the rest, there is no change and we get the distribution of the 43 corresponding codons

- (3+2) quartets  $(3+2) \times (3+1) = 20$
- 9 doublets  $9 \times 2 = 18$
- 2 singlets  $2 \times 1 = 2$
- 1 triplet  $1 + 2 = 3$

In Section 2, we mentioned the identity " $-384+384=0$ " as useful in applications. This is one good place, before

closing this Section, to show that it could help to establish the correct nucleon-number pattern in the 23 AAS-precursors (recall  $23 \text{ AASs} = 17 \text{ aas} + (S^{IV,II}, L^{IV,II}, R^{IV,II})$ ). Rakočević found the following results for the nucleon number of amino acid precursors, in their "head", "body" and "tail" (we reproduce his Table):

**Table 3:** The distribution of nucleon numbers of amino acid precursors between "head", "body" and "tail" (from Rakočević, 1998). *Symbols:* p-nfcaa: precursors of non four codon amino acids; p-fcaa: precursors of four codon amino acids.

|        | p-nfcaa     | p-fcaa     | S   | S'  | S''         |
|--------|-------------|------------|-----|-----|-------------|
| "head" | <b>746</b>  | 360        | 386 |     |             |
|        |             |            |     | 295 |             |
|        |             |            |     |     | <b>211</b>  |
| "body" | 909         | <b>228</b> | 681 |     |             |
|        |             |            |     | 84  |             |
| "tail" | <b>1237</b> | 472        | 765 |     |             |
|        | <b>2211</b> | 1741       |     |     | <b>2211</b> |

Rakočević, found that the (positive) Differences and the (positive) Differences of Differences is exactly 2211. Also, by making "crossings", or "zig-zags" i.e., numbers in bold, on the one hand, and the others, on the other hand, in columns 2 and 3, he obtained  $746+228+1237=2211$  and  $360+909+472=1741$ . Now, the sum of the nucleons in non-four-codon precursors ("head"+"body"+"tail") is equal to 2892 (no crossing) and the number of nucleons in four-codon precursors is equal to 1060. We have seen in Section 2 that the case (iii) for Eq.(7) corresponds to the total number of nucleons in 23 amino acid signal precursors 3952. It was obtained as  $3276+5 \times 128+36$  where 3276 is the sum of the divisors of 1989. First, writing  $3276+676$  and introducing the above mentioned identity " $-384+384$ ", we get  $2892+1060$  which is just the above partition of nucleon-numbers between the four-codon set and the non-four codon set (see Section 1). In the following Section, where water will be introduced, briefly, as a possible "player", we shall show that the "zig-zag" pattern  $2211+1741$ , found by Rakočević (see above), could be easily recovered.

#### 4. The "epiphany" of serine and an "intrusion" of water in the study of the genetic code

In the course of our most recent investigations (Négadi, 2011a), we have met several times some particular numbers that appear to fit those of some (experimentally) well known protonated serine clusters such as  $\text{Ser}^+$  (mw=106), the protonated monomer,  $\text{Ser}_2^+$  (mw=211), the protonated dimer and  $\text{Ser}_8^+$  (mw=841), the *protonated serine octamer*. These clusters of serine, especially the protonated serine octamer, are thought to have had some major role in the primordial biochemistry and homochirogenesis, at the origin of life (see for example Cooks *et al.*, 2001 and Hodyss *et al.*, 2001). Moreover, we have also surmised, and we are not alone (see for example Mišić, 2011), a possible role of *water* in the study of the genetic code (Négadi, 2011a). In the rest of this last Section, we shall try to explain why these two important additions, i.e., serine clusters and water, appear not only connected nicely with the amino acid precursors, considered in the first three Sections, but can also constitute with these latter a larger and more informative set than the set of 20 amino acids alone. In this trial, we also find a link to the "Self-Similar-Number", or fractal approach to the genetic code by Mišić (Mišić, 2010, 2011).

In Section 3, we have used some few successive (positive) difference operations, to analyze the chemical content of the amino acid precursors, which were first studied by Rakočević (Rakočević, 1998). Here, in this last and long Section we shall use a simplified version of the language of the Finite Difference Calculus applied to number sequences, mentioned in the introduction. Let S a given sequence  $s_0, s_1, s_2, \dots$ , S' the sequence of (positive) differences, S'' the sequence of (positive) differences of differences, and so on till the end of the process (no more difference to take) if the sequence is finite as it will be the case here. In Difference Calculus, one associates to the successive differences S', S'', ... the linear operator  $\Delta$  acting iteratively in the vector space of sequences ( $S' = \Delta S$ ,  $S'' = \Delta^2 S$ ,  $S''' = \Delta^3 S$ , ...). The link with the notation in Table 3 must be understood: the sequence S in the Table (386, 681, 765) is in fact itself a *pair-wise* difference between the sequence of nucleon-numbers in the non-four codon-set ("head"/"body"/"tail") and the sequence of nucleon-numbers in the four-codon-set

("head"/"body"/"tail"). The same applies also to Table 1, for atom-numbers. In this way and for the nucleon-numbers, Rakočević's pattern, 2211, corresponds to the sum of the numbers in the sequence  $S_{nuc}$  {386, 681, 765} of pair-wise differences and those in the sequence  $S_{nuc}' = \Delta S_{nuc}$ : {386, 84}.

It is interesting to note that an (overlooked) pattern, 211, of the same "kind" as the one above, would correspond here to the next iteration  $S_{nuc}'' = \Delta^2 S_{nuc}$  {211} (=295-84). This later, which closes the process (as there is no more difference to take), appears very interesting, in the following. As a first application, we now show that the sequence(s) for carbon-atom numbers in Eqs.(13)-(15) is very informative: its numbers will be shown to encode the detailed nucleon-number composition of the 23 amino acid precursors in Table 3. Let us look, in this vein, at the following tables

|         |     |    |     |          |    |    |     |     |    |    |     |    |
|---------|-----|----|-----|----------|----|----|-----|-----|----|----|-----|----|
|         |     |    |     | 23       |    |    |     | 6   |    |    |     | 61 |
| S       | 99  | 27 | 249 | S        | 27 | 99 | 249 | S   | 27 | 99 | 249 |    |
| S'      | 72  |    | 222 | S'       | 72 |    | 150 | S'' | 78 |    | 150 |    |
| S''     | 150 |    |     | S''      | 78 |    |     |     |    |    |     |    |
| Table I |     |    |     | Table II |    |    |     |     |    |    |     |    |

in two possible and interesting dispositions (I and II, see below). The numbers in the first row (I) correspond respectively to 23 AASs, 6 precursors and 61 amino acid precursors (same for II). Below is the corresponding sequence S for carbon-atom numbers. The third and fourth row gives the sequence of (positive) differences, S', and (positive) differences of differences, S''. We have immediately that all the numbers in I and in II add-up to respectively  $\Sigma^I=909$  and  $\Sigma^{II}=765$  and the sum of the numbers in the last rows (I and II) is  $s=228$ .

The means of the numbers in the third and fourth rows in each set have values  $m^I=148$  and  $m^{II}=100$ . Also the mean of the numbers in the third row in II is  $m=111$ , which is also the number of atoms in 23 amino acid precursors (p-fcaa, see Table 1). Now, the combination  $(\Sigma^I+s+m^{II})-\Sigma^{II}$  rewrites  $1237-472=765$  and we have also  $\Sigma^I-s=909-228=681$ . These last two relations are exactly the "body" and "tail" parts of Table 3. We have<sup>4</sup>  $2 \times 249 + m^I + m^{II} = 746$  and  $249 + m = 360$  so that the difference gives  $746 -$

$360=386$ . This is the remaining "head" part which completes our demonstration (see Table 3). As a second application, we use again the formalism of the successive difference iterations and apply it to the (remarkable) number 12935, considered above. We first show that it is related to the prime "quantum 037" and, second, it is also linked to serine. Consider for example the sum of all the numbers in the possible successive four iterations of the sequence of differences of S: {1, 2, 9, 3, 5}. We get

$$\sum_{i=1}^{i=4} \Delta^i S = 37 \quad (23)$$

where  $\Delta^i$  are the successive differences mentioned above. We see therefore that the prime number 37 is associated to the number 12935 describing the five standard genetic code multiplets (see the pattern in Eq.(19)) as well as "the degeneracies". We could also ask what are the values of Eq.(23) when permutations of the digits in the sequence S, are considered. For the 120 permutations, ranked by the index k, the sum (23) gives 37 in only two cases:  $k=1$ , computed above, and  $k=120$  for the "mirror" number 53921. For all 120 permutations, the sum varies between 13 and 41. Looking at (18) and (19), and also taking into account the remark following Eq.(19), we have, only four pairs of cases left, including "mirrors":

- (i) 12935: ( $k=1, 37$ ), 12953: ( $k=2, 31$ )
- (ii) 53912: ( $k=119, 40$ ), 53921: ( $k=120, 37$ )
- (iii) 21935: ( $k=25, 40$ ), 21953: ( $k=26, 34$ )
- (iv) 35912: ( $k=95, 34$ ), 35921: ( $k=96, 31$ ).

It appears, first, that the sum for the three cases  $k=1, 2, 120$ , is equal to  $31+2 \times 37=31+74=105$ . This is the nucleon-number in serine (side-chain and block). It is interesting to note that if the 120 cases are sorted in increasing values (from 13 to 41) then the two sum values "37" have, as new ranks, 105 and 106. These numbers are nothing but the nucleon numbers in serine

<sup>4</sup> The number 249 appears two times, first as  $27+222$ , in I and, second, as  $99+150$ , in II.

and its protonated monomer Ser<sup>+</sup> (105+1, see above). Also, the total sum of the k-values and the values of Eq.(23) for all the cases (i)-(iv) gives 768 or, 2×384, a value already met (see Section 2 and the remark after Eq.(8) about the number 384). The mean of the four pairs equal 192, the total number of nucleobases in the 64 codons.

Let us now return to the numbers (or "orbitals") 2211 and 211, for nucleon-numbers, 222+222, for carbon-numbers (see Eqs.(13)-(15)), and 222 or 224, for atom-numbers, depending on whether the "borrowing" is made or not (see the introduction). (Note that the last iteration Δ<sup>2</sup>S<sub>at</sub> is equal to 11 in the former case and 17 in the latter.) It is instructive to also note that these two numbers, 222 and 224, could be shown to encode remarkably the "tail"-part of atom-numbers of the amino acid precursors (see Table 1):

$$\frac{1}{2} [B_o(222+224)] - [a_o(222)+a_o(224)] = 138-59 = 79 \quad (24)$$

where we use the arithmetic functions a<sub>o</sub>, the sum of the prime factors, A<sub>o</sub>, the sum of the prime factors and their indices and finally B<sub>o</sub>, the sum of the prime factors their indices and their number<sup>5</sup> (see the Appendix in Négadi, 2011b). Also, we have the differences 224 - { $\frac{1}{2} [B_o(222+224)]$ } = 86 and 222 - { $\frac{1}{2} [B_o(222+224)]$ } = 84, which give the two possible values for the "body"-part in the non-four-codon set without and with "borrowing", respectively. Now, the sum of all the above numbers, in bold, gives 3088, using 222, and 3090, using 224. We have

$$B_o(3088) + B_o(3090) = 404 = 403 + 1 \quad (25)$$

The result in Eq.(25) is very interesting because, first, it is only one unit away from the total number of atoms in the 23 amino acid precursors in Table 1, 292+111=403 and, second, and on the one hand, the sum of its prime factors gives the (integer) molecular weight of serine, side-chain and block (Eq.(26))

$$a_o(404) = 105 \quad (26)$$

$$SPI(404) + \Omega(404) = 31 \quad (26)'$$

while, on the other, the sum of the prime-indices and the number of prime factors (Ω-function) gives the side-chain part (Eq.(26)') and, by subtracting from (26), the block part, 74, is recovered. Note also that B<sub>o</sub>(404)=105+31=136. It is also not difficult to show that the sum of (i) the numbers 404 and 136, (ii) their mean, as well as (iii) the B<sub>o</sub>-function of this latter gives 841, i.e., the number of nucleons (mw) of the protonated serine octamer Ser<sub>8</sub><sup>+</sup> (see at the beginning of this Section).

The number 3088, used in Eq.(25), is twice interesting. We have first that its B<sub>o</sub>-function gives the number of carbon atoms in 61 amino acid precursors, 249 (see Section 1) and, second, it is also related to Ser<sub>8</sub><sup>+</sup> because A<sub>o</sub>(3088)=249 and this is the protonated side-chain-part, 248+1 (8×31+1). The "block"-part is simply 841-249=592=8×74. Using again the number 249, we get B<sub>o</sub>(249)=113 which corresponds again to Ser<sub>8</sub><sup>+</sup> but in atom-number (8×14+1). The "alchemy" of the above numbers is such that it is literally possible to "see" the structure of serine, and more. First, 404+a<sub>o</sub>(404)=509.

Rakočević (1998), quoting Doolittle, says that the twenty amino acids have a "head" (the block) and a "tail" (the side-chain) but, for the precursors it is necessary, and fruitful, to add a third part, the "body", inexistent for an amino acid, in particular serine. It is as if this pattern was encoded in the sum of the digits of the above number 5+0+9=14, and serine has 5 atoms in its side-chain and 9 atoms in its block, and of course no "body" between them (see more below). More will be said on the number 509 below and it is interesting now to see that, concerning serine some higher order known species as (Ser<sub>8n</sub>+nH)<sup>+</sup> with n=1, 2 and 3, are easily derivable from our data in the Tables I and II, and also from the results obtained from the (very informative) "mixture" of carbon-, atom- and nucleon-numbers, considered above. We have seen above (see footnote 3) that the number 249 appears two times, or two times (Ser<sub>8</sub>+H)<sup>+</sup>, from carbon-atom numbers. We have also

<sup>5</sup> It is clear that the factor ½ could be thought as coming from computing the mean of, on the one hand, the sum of the prime factors, 225 and, on the other, the sum of the prime indices and the number of prime factors, 51.

found this same number 249, *one time*, as  $A_0(3088)$  above.

Now, *two times* 249 fits the case  $n=2$  ( $16 \times 31 + 2$ ) and *three times* 249 the case  $n=3$  ( $24 \times 31 + 3$ ). This latter case could also be found by taking the mean of the sums in the tables I and II of all the numbers (in S, S' and S''):  $(819+675)/2=747=24 \times 31 + 3$ . The nice thing with the number 249, as mentioned above, is that its  $B_0$ -function gives  $B_0(249)=113$  which is nothing but, once again,  $Ser_8^+$  the *protonated serine octamer* but here in *atom-numbers*: serine has 14 atoms so that  $14 \times 8 + 1 = 113$  and we remind that the 14 atoms of serine<sup>6</sup> were found above as  $5+0+9$ . It is not difficult to see that the reasoning made above for nucleon-numbers in the side-chains and leading to the series  $(Ser_{8n} + nH)^{+n}$  with  $n=1, 2$  and  $3$  remains valid, in atom-numbers:  $n \times B_0(249)$ . As a remark, the sum  $m^I + m^{II} = 248$  obtained from the tables I and II (and also 840 when considering the blocks) corresponds to the (still not found but predicted) neutral serine octamer  $Ser_8^0$ . For the smaller serine clusters like  $Ser_2^+$  ( $mw=2 \times 105 + 1$ ) we have already the number 211 for the difference of difference in nucleon-numbers ( $=\Delta^2 S_{nuc}$ , see Table 3).

The numbers **2211** and **211**, and also the two (possible) values for atom-numbers 222 and 224, could be again used to derive several other interesting results. First, the number of coprimes to the sum  $2211+211$ , or  $\phi(2211+211)$ , gives 1032 which fits nicely the total number of nucleons in the eight components of DNA and RNA<sup>7</sup>. Now, from Eqs.(25)-(26)

$$B_0(3088) + B_0(3090) + a_0(404) = 404 + 105 = 509 \quad (27)$$

which fits the number of nucleons in the four RNA-components and, by subtracting from

<sup>6</sup> The result in Eq.(27),  $509=404+105$ , tells more by taking the sum of the digits: the block has **4** hydrogen, **2** carbon, **2** oxygen and **1** nitrogen atoms. As for the side-chain part, it is in fact serine's octamer, "from the heavens", that could tell us the details: remember, we have shown in the text the relation  $B_0(249)=113$ , which links the nucleon-number and the atom-number of the protonated serine octamer. The sum of the digits in the last number gives the answer: 1 carbon, 1 oxygen and 3 hydrogens.

<sup>7</sup> The number of nucleons and the number of atoms in the nucleobases are the following DNA: C (111, 13), T (126, 15), A (135, 15), G (151, 16); RNA: C (111, 13), U (112, 12), A (135, 15), G (151, 16).

the  $\phi$ -function above, we get the DNA-component  $1032-509=523$  (see the footnote 5 below. Thus, the number 509, as mentioned above, has been very useful. Second, using the same  $\phi$ -function and adding the mean value of the other two numbers 222 and 224, gives  $\phi(2211+211) + \frac{1}{2}(222+224) = 1255$  and this is the number of nucleons in the 20 amino acids side-chains, an important number. Note that this same number could be obtained, otherwise, by taking the sum of the nucleons in the precursors (see Table 2) and the number of nucleons in serine, from Eq.(26) above:  $1150+105=1255$ . Finally, taking the  $\phi$ -function of the sum *and* the difference of 2211 and 211, as well as the one for the above obtained number, 1255, leads to  $1032+800+1000=2832$ . This number fits the total number of *atoms* in the 64 DNA codons ( $48 \times 59$ , see the footnote 5). Let us now take the sum  $2211+224+222+Eq.(26)'$ , where only nucleon and atom-number contributions are considered and Eq.(26)' is a "contribution" from serine, via the number of nucleons in its side-chain, 31. The result of this sum, the number 2688, could also be written  $48 \times 56$ , or  $48 \times (12+13+15+16)$  which is nothing but the right number of atoms in the 64 RNA-codons. As a supporting supplementary result, we have that the  $B_0$ -function of the sum of the number of atoms for DNA and RNA gives  $B_0(2832+2688)=64$ , i.e., the total number of codons and, by elementary manipulations of the terms of the prime factorization ( $2^4 \times 3 \times 5 \times 23$ ) we could write

$$23+38+3=61+3 \quad (28)$$

where 23 is for the twenty-three AASs or, 17  $aas + S^{IV,II} + L^{IV,II} + R^{IV,II}$ , and the degeneracies for the multiplets are as follows: quartets  $8+3+4=15$ , sextets  $3+(4+5)=3+9=12$ , doublets 9, triplet 2 (with total 38) and finally "stops" 3.

It is interesting to note that the two numbers 3088 and 3090, mentioned and used above, have also an interesting mean and applications. The  $B_0$ -function of their *mean* gives

$$B_0\left[\frac{1}{2}(3088 + 3090)\right] = 3532 \quad (29)$$

which compares well with the result in Eq.(6), giving the number of nucleons in 20 amino acid precursors. To the first of these two numbers, we could add the  $B_0$ -function of our starting relation in Eq.(1)

$$3088+B_0(2211-222)=3088+57=3145 \quad (30)$$

The number in Eq.(30) corresponds exactly to the number of nucleons in the 23 AASs. Now, the sum leading to 3088 could, at will, be partitioned into 2211, on the one hand, and 444+222+211, on the other. Next we subtract the quantity  $\bar{\Psi}_{(0,1,0,0)}(\Psi) + \bar{\Psi}_{(0,0,3,0)}(\Psi)=768$  (see Eq.(4)) from the first number and add it to the second and, finally putting everything in Eq.(30) we end up with 1443+1702. The first term, 1443, corresponds to the number of nucleons in the 23 AASs side-chains and 1702 to the 23 AASs blocks (23×74). Finally, the set of 23 amino acids signals, itself, as 20+3, is easily derivable from the above results, thanks to serine from Eq.(26). We have 2211-768=1443, the nucleon-number in the 23 AASs, 1150+105=1255, the nucleon-number in the 20 amino acids and, by subtracting 1443-1255=188, the nucleon-number in the three sextets serine, leucine and arginine.

Another way to get the above number of nucleons in the 23 AASs, consists in using the other number, 3090, and add the total number of digits for all the numbers, 19, and sum of their values, 36 (for carbon write 222+222 and, for atom-numbers take 222 and 224)<sup>8</sup>. We get 3090+55=3145. (Note, interestingly, that the sum of the nucleons in the six precursors (Table 2), their number and  $\Psi=2211-222$  from our starting Eq.(1) leads to this same number (1150+6+1989=3145). We have seen, in this Section, that not only serine "shows itself" several times but it also "contributes" concretely to get meaningful results and it appears that this "phenomenon" is not exhausted: it has been experienced recently using a completely different approach (Négadi, 2011a). Let us call again the number 1152, see Eqs.(1) and (2), to which we promised to return here.

We have, first, that  $A_0(1152)=31$ , the nucleon number in serine's side-chain.

Second,  $a_0(\sigma(1152)+\phi(1152)+1152)=31$ , where  $\sigma$  is the sum of the divisors-function. We have therefore the same number, again. We already know that  $\phi^{(2)}(\Psi)=384$ ,  $\phi^{(3)}(\Psi)=128$ , could be interpretable as respectively the number of atoms in the 20 amino acids (side-chain and block) and the total number of codons in DNA and RNA (2×64). Concerning precisely this last point, we have that the product of the number of precursors, 6, with the number 1152 gives 6×1152=6912 and, remembering that the six precursors are frequently grouped into two sub-sets according to the pattern 3+3 (see Rakočević, 2007 and Négadi, 2011b) we get 3456+3456=6912. This is the pattern for atom-numbers in the 64 RNA-codons (192 nucleobases) in the Sukhodolets Genetic Code Table (Cf. Rakočević, 2007 ): 3456 atoms in two inner and 3456 atoms in two outer columns of the standard genetic code. Note that, here, the *ribonucleosides* UMP (34), CMP (35), AMP (37) and GMP (38) have been used. (The number of atoms in these latter are obtained from those of U, C, A and G by adding 22 atoms, see footnote 7.) It is interesting to note that the above number 1152 leads to the same result as above but with a shift:  $\phi(1152)=384$ ,  $\phi^{(2)}(1152)=128$  ( $=2\times\phi^{(3)}(1152)$ ). Another promised remark about the protonated serine octamer and the important number 853 (see Négadi, 2011b) is worth to mention here. It appears that, in recent experiments, the "center mass" of the (Tandem Mass Spectrometry) distribution of serine octamers is (865+841)/2=853. Recall that we have found this number, in Section 2, as a mean, in connection with serine's octamer. Serine, as explained above, is a unique amino acid, in connection to homochirality. We have also seen above that many of its (numeric) facets, and also those of its clusters, are encoded in tables I and II.

As a matter of fact, taking the sum of all numbers (including the first rows and repetitions) one would find 1674 and, one step more, obtain  $B_0(1674)=65$ . This latter number has as prime factorization  $2\times3^3\times31$  so that, letting "serine's number 31" act, through it prime-index 11, one could put the final result in the form 11+54=65. Note that 1674/31=54). The relation 65=11+54 seems to give more status to the results about

<sup>8</sup> 2211, 211, (222+222), (222+224): 19 digits and total value 36.

serine, as implied in homochirogenesis, because there are exactly 65 Chiral Space Groups (Sohnke space groups) including 11 enantiomeric pairs. This same relation has also been obtained recently, in the context of the irregular tetrahedron classification model of the 20 amino acids (Négadi, 2011b).

We end this Section with two remarks, one of them concerns the prime number 37, which was found, here (see Eq.(23), tightly bound to the number 12935 (and to its "mirror" 53921) and, the other connected with water. First, we have by combining the number 2211 (*differences and Difference of Differences in nucleon-number in 23 amino acid precursors*), which was so much used in the above Sections, and also the number of nucleons in the fundamental set of the six precursors as well as their number (see Table 2). We get

$$2211+(1150+6)=3367 \quad (31)$$

Note that the sum of the two terms in the parenthesis is also recovered as  $\varphi(\Psi)+\Omega(\Psi)=1152+4=1156$ , see Section 2. We know from our exchanges with Mišić [also, about the Interface of Classical realm and Quantum realm in the living organisms more in this Issue of NeuroQuantology: (Mišić, 2011)] that this latter number is first a multiple of 37 and second it is the second term of a series she calls "decimal varieties of 37"; the third term being 333667. Note that 3367 is a multiple of 37 but 333667 (28693<sup>th</sup> prime) is not; it is however very nearly divisible by 37 and, using the  $\varphi$ -function, we have  $\varphi(333667)/37=9018$  (see below, for an application).

It is interesting to note that, from our own results in the above Sections, the *sum of the two first members*, 37 in Eq.(23) and 3367 in Eq.(31) gives 3404, the number of nucleons in the 61 coded amino acids side-chains (Eq.(5), see for example Négadi, 2009). Also, as  $3367=91 \times 37$ , we could take the sum of these two figures (with sum  $2 \times 64$ ) and add it to the preceding result to get  $3367+37+(37+91)=3532$  which appears to be equal to the total number of nucleons in 20 amino acid precursors, in Eq.(6). Note, interestingly, that  $A_0(3404)=27+37=64$  and  $B_0(3404)=64+27=91$ , that is just  $3367/37$ . The third member 333667 is also interesting

because  $B_0(B_0(333667))=204$  and this latter writes also, and willingly,  $66+138$ . We have thus found the number of atoms in the 20 amino acids side-chains with 66 atoms in the quartets and the sextets, on the one hand, and 138 in the doublets, the triplet and two singlets, on the other.

Another application of the number 333667 (promised above) concerns the genetic code's number 12935 found in Section 3, Eqs.(18) and (19), and its particular permutation 21953 which could be easily shown to leave *invariant* the multiplet structure and the degeneracy. We have  $12935+\varphi(333667)/37=21953$ . Our second remark concerning *water* is, in fact, connected not only with the above "037 varieties" but also with *serine* and, consequently *chirality*. A whole paper will be devoted to these "connections" and, here, we give only some few preliminary elements. Our recent paper devoted to the integer (Heronian) tetrahedron classification of the 20 amino acids (Négadi, 2011a) relies on an important number, the tetrahedron (integer) volume  $V=18144$ . It appears that the number of numbers coprime to this latter number and to 3367, or their  $\varphi$ -functions, are connected:  $\varphi(V)=2 \times \varphi(3367)$ .

It is an established fact that water has unique properties and these are connected with *hydrogen-bonding*. This latter, on the other hand, is crucial to life and its whole biochemistry. To see the "connection" one has to consider for example the "body-centered cubic lattice-gas model of water" by Besseling and Lyklema (1994) which, it has been shown, "captures the essential physics of water". Without entering into the details, these authors, in their statistical treatment, began by evaluating the number of contact-types and their "degeneracy" (the number of ways to realize a certain type of contact). For two neighboring water molecules there are 144 configurations and only in 18 (out of 144) a hydrogen-bond is formed between them and there are therefore  $144-18=126$  configurations that do not lead to a hydrogen-bond.

It is at this point where we have present three ways, at least, to "see" the above mentioned connection and the first comes from the volume  $V=18144$  of the integer tetrahedron which could tell us something interesting related to the above

"degeneracies". In base-1000, the "digits" are "18" and "144". In base-100, the "digits" are "1", "81" and "44" with sum 126. Finally, in ordinary base-10 the sum of the digits is equal to  $18=(1+8)+(1+4+4)=(9+9)^9$ . Our second way consists in considering again our two numbers 12935 and 21953. It appears that their sum, with prime decomposition  $2^3 \times 7^2 \times 89$ , leads to  $A_0(12935+21953)=144$ , the total number of "contacts". As 89 is the 24<sup>th</sup> prime, we could either write  $144=31+113$  or  $18+126$  where, in the first case we used  $89+24=113$  and, in the second, we used  $89+24+(6+7)=126$ . (In the first case, one would see the "hand" of serine (31) and its protonated octamer (113), see above.) The third way comes entirely from carbon. As a matter of fact, from Section 3, we know that, the sum of differences and difference of differences (for carbon) gives  $222+222=444$  and we have  $\phi(444)=144$ . Also, the sum of carbon atoms in the 6 aa-precursors (27) and in 23 AASs (99) gives 126. These two numbers, 126 and 144 are both successive divisors (out of 60), the 30<sup>th</sup> and 31<sup>th</sup> with rank-sum 61, of the number 18144. From Eqs.(25)-(26)' and water "numbers", we have

$$404+B_0(404)+(126+144) \quad (32)$$

$$+B_0(126+144)=249+592=841$$

where we used  $B_0(3088)=249$ , in Eq.(25). The above relation fits perfectly the integer molecular weight of the protonated serine octamer according now to the precise partition of nucleon numbers between the side-chains (249) and the blocks ( $592=8 \times 74$ ). Finally,  $B_0(18144)+B_0(126+144)=56+31=87$ . This is the molecular weight of serine's "residue", that is with a block having one water molecule less ( $105-18=87$ ). Is water ( $H_2O$ ) really encoded in the volume of the integer tetrahedron:  $1+8+1+4+4=1+1+(8+8)=1+1+16?$

## References

- Besseling NAM and Lyklema J. Equilibrium Properties of Water and Its Liquid-Vapor Interface. *Phys Chem* 1994; 98: 11610-11622.
- Cooks RG, Zhang D, Koch KJ, Gozzo FC and Eberlin MN. Chiroselective self-directed octamerization of serine: implications for homochirogenesis. *Anal Chem* 2001; 73: 3646-3655.
- Hodyss R, Julian RR and Beauchamp JL. Spontaneous chiral separation in noncovalent Molecular clusters. *Chirality* 2001; 13: 703.
- Mišić NŽ. The Self-similar Numbers as a Special Case of Cyclic Numbers and Their Relation to the Cyclic (Genetic) Codes. *Proceeding of Neurel* 2010; 97-102.
- Mišić NŽ. Nested Numeric/Geometric/Arithmetic Properties of shCherbak's Prime Quantum 037 as a Base of (Biological) Coding/Computing. *NeuroQuantology* 2011; 9(4) [*this Issue*].
- Négadi T. The genetic code via Gödel encoding. *The Open Physical Chemistry Journal* 2008; 2: 1-5.
- Négadi T. The genetic code degeneracy and the amino acids chemical composition are connected. *NeuroQuantology* 2009; 7(1): 181-187.
- Négadi T. On Rakočević's amino acid biosynthetic precursors relations. *NeuroQuantology*, 2011b; 9(4) [*this issue*].
- Négadi T. The irregular (integer) tetrahedron as a warehouse of biological information. *Sumited to Symmetry: Culture and Science*, Nov 2011a.
- Rakočević MM. Genetic Code Table: A note on the three splitting into amino acid classes. arXiv:0903.4110v1 [q-bio.BM], 2009.
- Rakočević MM and Jokić A. Four stereochemical types of protein amino acids: synchronic determination with chemical characteristics, atom and nucleon number. *J Theor Biol* 1996; 183: 345-349.
- Rakočević MM. Private communication (August-December 2011).
- Rakočević MM. Whole-number Relations Between Protein Amino Acids and Their Biosynthetic Precursors. *J Theor Biol* 1998; 191: 463-465.
- Rakočević MM. Genetic Code Table: Four diversity types of protein amino acid classes. arXiv:1107.1998v2[q-bio.OT] 2007.
- Rumer Y. About systematization of the genetic code. *Dok. Akad Nauk SSSR* 1966; 167: 1393-1394.
- Stryer L, Berg JM and Tymoczko JL. *Biochimie. Médecine-Science*, 2003, Flammarion.
- shCherbak VI. Arithmetic inside the universal genetic code. *BioSystems* 2003; 70: 187-209.
- shCherbak V. The Arithmetical Origin of the Genetic Code. In *The Codes of Life: The Rules of Macroevolution*, M. Barbieri, Ed. Springer, 2008; 153-188.

<sup>9</sup> Dr. Klass Besseling, from the Department of Chemical Engineering, Delft University of Technology, Netherlands, is warmly acknowledged for very clear and kind explanations on the computation of the degeneracies.