Article

Revealing the genetic code symmetries through computations involving Fibonacci-like sequences and their properties

Tidjani Négadi

Physics Department, Faculty of Exact and Applied Science, University of Oran, Oran1, Algeria; tnegadi@gmail.com

Abstract: In this work, we present a new way to study the genetic code mathematical and chemical structure, based on the use of mathematical computations involving some few recently designed Fibonacci-like sequences, the "seeds" ("initial conditions") of which are chosen according to the chemical and physical data of the three amino acids serine, arginine and leucine, playing a prominent role in a recent symmetry classification scheme of the genetic code. It appears that these mathematical sequences, of the same kind as the famous Fibonacci series, apart from their usual recurrence relations, are highly interwinned by many useful linear relationships. Using these sequences and also various sums or linear combinations of them, we derive several physical and chemical quantities of interest, as the number of total coding codons, 61, obeying various degeneracy patterns, the detailed number of hydrogen atoms, the detailed number of atoms (H/CNOS) and the integer molecular mass (or nucleon number), in the side chains of the 61 "amino acids", also in various degeneracy patterns, in agreement with those described in the literature. Also, unexpectedly but interestingly, we find, as a by-product, an accurate description of the very chemical structure of the four ribonucleotides uridine monophosphate (UMP), cytidine monophosphate (CMP), adenosine monophosphate (AMP) and guanosine monophosphate (GMP), the building blocks of RNA whose groupings, in three units, constitutes the triplet codons. In summary, we find a full mathematical and chemical connection with the "ideal sextet's classification scheme" mentioned above as well as with several others, notably, the Findley-Findley-McGlynn symmetrical classification. We organize the content of the text in such a way that, besides the presentation of several new research results, it has also an educational dimension. The paper could therefore be read and the computations easily worked out, also by non-experts with mathematical background.

Keywords: genetic code; symmetries; Fibonacci-like sequences; amino acids; ribonucleotides; patterns; hydrogen, atom; molecular mass

1. Introduction

In this paper, we present a new way to study the genetic code and its symmetries, using mathematics, more precisely, a small set of brand new Fibonacci-like sequences and, occasionally, some (useful) well known elementary functions from number theory. Using these sequences, the whole and detailed chemical content of the set of amino acids, as structured by several well known symmetry patterns, including their degeneracy, is revealed. Also, several other original applications, using the above sequences, are carried out. Before presenting all these results, let us begin by giving a brief summary of the genetic code. This latter is the basis of life on Earth and was masterfully deciphered in the mid-sixties of the last century [1]. It is the great biological "dictionary" that translates the language of DNA/RNA, which transmit the inherited information located in the genes, to the language of proteins which carry out the biological constructions and functions. It is well known that the "alphabet" of the former language consists of four fundamental units, the nitrogenous bases T (thymine), C (cytosine), A (adenine) and G (guanine) for DNA and U (uracil), C, A and G for RNA. As for the "alphabet" of the second language, it comprises the set of 20 amino acids. In the process of translation between these two languages, in the ribosome for short, there are $64 = 4^3$ "words", the codons, which are

made of three nucleotides. In the standard genetic code, only 61 of such codons are translated into amino acids and three others serve as termination or stop codons. To a given codon, from the set of 61, corresponds one and only one amino acid. Now, it is well known that several codons could code for the same amino acid so that the genetic code is said degenerate and the 20 amino acids are organized into "multiplets". These are as follows: 5 *quartets*, each, coded by four codons proline (P, Pro), alanine (A, Ala), threonine (T, Thr), valine (V, Val) and glycine (G, Gly); 9 *doublets* each coded by two codons phenylalanine (F, Phe), tyrosine (Y, Tyr), cysteine (C, Cys), histidine (H, His), glutamine (Q, Gln), glutamic acid (E, Glu) and aspartic acid (D, Asp), asparagines (N, Asn) and lysine (K, Lys); 3 *sextets* each coded by 6 codons serine (S, Ser), arginine (R, Arg) and leucine (L, Leu); 1 *triplet* coded by three codons isoleucine (I, Ile) and finally 2 *singlets* each coded by one codon methionine (M, Met) and tryptophane (W, Trp). In the parentheses, the one-letter and three-letter codes of the amino acids are indicated. Below, in Table 1, we show the genetic code table (the three stop codons are indicated in green color background).

Table 1. The genetic code table (format from Négadi, [2])

UUU (F)	UUC (F)	**UCU (S)**	**UCC (S)**	**CUU (L)**	**CUC (L)**	CCU (P)	CCC (P)
UUA (L)	**UUG (L)**	**UCA (S)**	**UCG (S)**	**CUA (L)**	**CUG (L)**	CCA (P)	CCG (P)
UAU (Y)	UAC (Y)	UGU (C)	UGC (C)	CAU (H)	CAC (H)	**CGU (R)**	**CGC (R)**
UAA (STP)	UAG (STP)	UGA (STP)	UGG (W)	CAA (Q)	CAG (Q)	**CGA (R)**	**CGG (R)**
AUU (I)	AUC (I)	ACU (T)	ACC (T)	GUU (V)	GUC (V)	GCU (A)	GCC (A)
AUA (I)	AUG (M)	ACA (T)	ACG (T)	GUA (V)	GUG (V)	GCA (A)	GCG (A)
AAU (N)	AAC (N)	**AGU (S)**	**AGC (S)**	GAU (D)	GAC (D)	GGU (G)	GGC (G)
AAA (K)	AAG (K)	**AGA (R)**	**AGG (R)**	GAA (E)	GAG (E)	GGA (G)	GGG (G)

In this work, the "anomalous" three amino acids serine, arginine and leucine, each coded by six codons, will play a prominent role. Contrary to all other 17 amino acids the codons of which share the same first base, the three mentioned amino acids have, each, their 6 codons distributed into two different *family boxes*. (There are 16 such family boxes and each one of them is a set of four codons having the same first and second base (see Table 1)). The structure of the three sextets is the following serine: {UCN, AGY}, arginine {CGN, AGR}, leucine {CUN, UUR} (N for any base, Y for pyrimidine U or C and R for purine A or G). There are more and more voices rising to underline or put emphasis on the singular nature of the three sextets and also to bring experimental data which tend to show it [3, 4]. On the other side, people have always taken for granted that the number of amino acids encoded by the (standard) genetic code is 20. Yes, this is true but this view is evolving. Some few years ago, a published work, [5], claimed that this number has to be increased to 23 by considering that the quartet part and the doublet part of each one of the 3 sextets as *distinct*. In this work, the authors present a new "*effective number of codons*", called N_c, to characterize codon usage bias in the analyzes of protein-coding genes, which improves existing ones. Here, N_c is shown to be a better predictor when its value is increased from 20 to 23 with, in particular, each 6-fold codon set (each sextet as it is called in this work) is considered to be composed of *separate* 4-fold and 2-fold parts. These six entities are Ser[II,IV], Arg[II,IV], Leu[II,IV] which, added to the 17 remaining amino acids with no "degeneracy" at the *first base position*, as mentioned above, give a total of 23. This number, together with the remaining degenerate codons, 38, constitutes the pattern "23 + 38" the importance of which not only has been established in our recent works (see [2] and the references therein) but is also again established all along the present work. One of these kind of approaches is particularly relevant to the present work, the "Ideal" symmetry classification scheme, introduced some few years ago in [6]. It will be summa-

rized in section 2.3 and we present its numerous links with the present work in section 4.2. In section 2, we summarize three important symmetries of the genetic code, including the above mentioned one. In section 3, we present our new Fibonacci-like sequences and their properties, which are the main mathematical tools used in this paper. In section 4, we apply these sequences to derive the degeneracy structure of the 61 amino acids, their hydrogen atom, atom and nucleon number content, as structured by the symmetries mentioned in section 2 and various other remarkable patterns. In section 5, still using some elements of our sequences, we make contact with the work by shCherbak, [7], concerning the singular structure of proline and derive, in a totally original way, the mathematical form of the shCherbak-Makukov "activation key, [8], which, as is well known, lead to many remarkable and beautiful nucleon number patterns comprising, in particular, those related to Rumer's symmetry (see below about this latter). In section 6, using the "seeds" of our Fibonacci-like sequences, that is their "initial conditions", and *only these*, we find that they are capable, on their own, to provide the very hydrogen atom content of the 61 amino acids, derived in the various patterns considered in section 4. We strongly recommend the reader, at this point, before going to the next sections and get a comfortable reading of them, to take a look at Appendix A, which summarizes the chemical data of all 20 amino acids, including also the degeneracies and Appendix B, where a few other mathematical tools, used in this paper, are defined with the visualization of some computation examples.

2. The symmetries of the genetic code

2.1. Rumer's symmetry

The oldest known symmetry of the genetic code has been discovered by Rumer in 1966, [9]. This symmetry, which is defined by the transformation $U \leftrightarrow G, A \leftrightarrow C$, divides the genetic code 8×8 table into two equal halves of 32 codons each, we call them here M_1 and M_2. The set M_1 (shown in grey background in Table 1) comprises 8 quartets of codons, each, having the same two first-bases and coding for the same amino acid, the third base being irrelevant. In this set, among the 8 quartets, 3 correspond to the quartet part of the 3 sextets serine, arginine and leucine. The set M_2 comprises group-I amino acids (2 singlets), group-II amino acids (9 doublets), group-III amino acid (1 triplet) and also 3 stops or termination codons (see Table 1). The point here, concerning symmetry, is that under Rumer's transformation, performed on all three bases, the sets M_1 and M_2 are exchanged: $M_1 \leftrightarrow M_2$.

2.2. The 3rd base symmetry classification

In 1982, Findley et al., [10], by viewing the genetic code as a relation, rather than a mapping, extracted a fundamental symmetry for the doubly degenerate codons (group-II). The 64-codons set is partitioned into four disjoint sub-sets and each one of them contains only codons having the *same third base* (see Table 2, below). These authors establish relations which define a *one-to-one* correspondence between one member of a *doubly degenerate* codon pair to the other member (see the reference above for details). These relations could be stated, in words, as follows: (i) if a codon for an amino acid has 3rd base U, then there is a codon for the same amino acid having 3rd base C and vice versa **OR** (ii) if a codon for an amino acid has 3rd base A, then there is a codon for the same amino acid having 3rd base G and vice versa. For a doubly degenerate codon pair (i) and (ii) are mutually exclusive. For order-4, or quartets, (i) and (ii) hold simultaneously. For order-6, the sextets, the quartet part obeys (i) AND (i) and, for the doublet part one has (i) OR (ii). For the odd-order degenerate codons (Ile, Met and Trp), however, there is a slight deviation from symmetry. Below, in Table 2, we show this classification

Table 2. The 3rd base classification of the 64 codons, [10].

UCU	Ser (6)	UCC	Ser (6)	UCA	Ser (3)	**UCG**	Ser (3)

AGU		AGC		AGA	Arg (20)	AGG	Arg (20)
CGU	Arg (10)	CGC	Arg (10)	CGA		CGG	
CUU	Leu (9)	CUC	Leu (9)	CUA	Leu (18)	CUG	Leu (18)
GCU	Ala (4)	GCC	Ala (4)	UUA		UUG	
GUU	Val (7)	GUC	Val (7)	GCA	Ala (3)	GCG	Ala (3)
CCU	Pro (5)	CCC	Pro (5)	GUA	Val (7)	GUG	Val (7)
GGU	Gly (1)	GGC	Gly (1)	CCA	Pro (5)	CCG	Pro (5)
ACU	Thr (5)	ACC	Thr (5)	GGA	Gly (1)	GGG	Gly (1)
UUU	Phe (7)	UUC	Phe (7)	ACA	Thr (5)	ACG	Thr (5)
UAU	Tyr (7)	UAC	Tyr (7)	CAA	Gln (6)	CAG	Gln (6)
UGU	Cys (3)	UGC	Cys (3)	AAA	Lys (10)	AAG	Lys (10)
CAU	His (5)	CAC	His (5)	GAA	Glu (5)	GAG	Glu (5)
GAU	Asp (3)	GAC	Asp (3)	UAA	STOP	UAG	**STOP**
AAU	Asn (4)	AAC	Asn (4)	UGA	STOP	UGG	Trp (8)
AUU	Ile (9)	AUC	Ile (9)	AUA	Ile (9)	AUG	Met (7)
Hydrogen	84		84		92		98
Nucleons		1728				1676	

2.3. The weak/strong, purine/pyrimidine and keto/amino symmetries

The main idea behind the "ideal" symmetry classification scheme by Rosandić and Parr, [6], mentioned above, is to consider the three sextets serine, arginine and leucine, each encoded by six codons, as "*initial generators*" with serine playing the central role. This scheme divides the 64 codons table into two groups of 32 codons each, the "leading" group and the "nonleading" group and each one of them consists of A+U rich and G+C rich (equal) parts. The "ideal" classification scheme is generated by combining the six codons of serine, arginine and leucine in the following manner: serine, the "initial" generator with its six codons, arginine also with its six codons and leucine with *only* the quartet part of its six codons part define the "leading" group (with 32 codons). The remaining doublet part of leucine, on the other hand, constitutes a "seed" for the construction of the "nonleading" group (with 32 codons). The whole set {Ser^{IV-II}, Arg^{IV-II}, Leu^{IV-II}} is called by the above authors the "core"; its members are underlined in Table 3 below.

Table 3. The "ideal" symmetry classification scheme, [6].

UUU (F)	UUC (F)	**UCU (S)**	**UCC (S)**	**CUU (L)**	**CUC (L)**	CCU (P)	CCC (P)
UUA (L)	**UUG (L)**	**UCA (S)**	**UCG (S)**	**CUA (L)**	**CUG (L)**	CCA (P)	CCG (P)
UAU (Y)	UAC (Y)	UGU (C)	UGC (C)	CAU (H)	CAC (H)	**CGU (R)**	**CGC (R)**
UAA (STP)	UAG (STP)	UGA (STP)	UGG (W)	CAA (Q)	CAG (Q)	**CGA (R)**	**CGG (R)**
AUU (I)	AUC (I)	ACU (T)	ACC (T)	GUU (V)	GUC (V)	GCU (A)	GCC (A)
AUA (I)	AUG (M)	ACA (T)	ACG (T)	GUA (V)	GUG (V)	GCA (A)	GCG (A)
AAU (N)	AAC (N)	**AGU (S)**	**AGC (S)**	GAU (D)	GAC (D)	GGU (G)	GGC (G)
AAA (K)	AAG (K)	**AGA (R)**	**AGG (R)**	GAA (E)	GAG (E)	GGA (G)	GGG (G)

As explained, at length, by the authors, the genetic code table in this new scheme is created by codons sextets based on exact *purine/pyrimidine* symmetries (YR: (U, C, A, G) → (C, U, G, A)), A+U-rich/C+G-rich symmetries, *strong/weak*, or complementary, symmetries (SW: (U, C, A, G) → (A, G, U, C)) and *keto/amino* symmetries (KM: (U, C, A, G) → (G, A, C, U)) (see [6]). By starting with serine, the *initial generator* with its 6 codons, the whole "leading" group (32 codons) is created using transformations among those mentioned above and some mapping rules. In an analogous manner, starting from the two codons of leucine, as "seeds", the whole "nonleading" group is constructed. There is also a simple relation between the "leading" group and the "nonleading" group (see the reference 6 mentioned above). We show, in Table 3, for visualization, these two groups by using our

own format of the genetic code table (the "leading" group is shown in grey background). It is also interesting to note that, under Rumer's transformation $U \leftrightarrow G, A \leftrightarrow C$,, the "leading" group remains *globally invariant* whether the transformation is applied to the first base , only, to the first two bases, only, or to all the three bases, and same for the "nonleading" group.

Below, in section 4.2, we will show, that the three amino acids serine, arginine and leucine, will also play a prominent role, as mathematical (and chemically inspired) "seeds" to compute the chemical content of the twenty amino acids, including degeneracy.

3. A rich set of Fibonacci-like sequences and their properties

Let us introduce now, as stated in the introduction, four Fibonacci-like sequences which will prove resourse-rich and very interesting in their applications throughout this work. (Another fifth sequence, just as interesting, will be introduced later, in Equ.(26).) They are also called (p, q)-Fibonacci sequences and are defined by the following common defining relation

$$pF_{n-1} + qF_{n-2}, \tag{1}$$

where F_n is the ordinary Fibonacci number. These four sequences differ only by the data of the numbers p and q which, here, play the role of "initial conditions" or "seeds", as we will call them throughout this paper. Below, in the next section, we shall explain and *justify* the choice of these "seeds" but, for the moment, we introduce the four mentioned sequences by giving a name to each one of them and their "seeds": (i) a_n: $p = 1, q = 6$, (ii) a'_n: $p = 6, q = 1$, (iii) b_n: $p = 9, q = 13$, (iv) c_n: $p = 5, q = 30$. In Table 4 below, we give the first few terms

Table 4. The first few terms of the Fibonacci-like sequences a_n, a'_n, b_n and c_n

n	1	2	3	4	5	6	7	8	9	10	11	12	13
a_n	6	1	7	8	15	**23**	**38**	**61**	99	160	259	419	678
a'_n	1	6	7	13	20	33	53	86	**139**	225	364	589	953
b_n	13	9	22	31	53	84	137	221	**358**	579	937	1516	2453
c_n	30	5	35	40	75	**115**	**190**	305	495	800	1295	2095	3390

These sequences obey several linear relations, some of which will prove very interesting in view of their applications in this work. They are presented below, in Equ.(2), and could be easily checked

$$
\begin{aligned}
&\text{(i)} & a_{n,} + b_{n+1} &= a_{n+4}, \\
&\text{(ii)} & a_{n,} + a_{n+6} &= 2b_{n+2}, \\
&\text{(iii)} & b_{n,} + b_{n+2} &= c_{n+2}, \\
&\text{(iv)} & b_n + c_{n+1} &= 2b_{n+1}, \\
&\text{(v)} & c_n + 2b_{n-1} &= b_{n+2}, \\
&\text{(vi)} & b_n + c_{n+3} &= b_{n+4}, \\
&\text{(vii)} & a_n + c_{n+3} &= 2a_{n+5}, \\
&\text{(viiii)} & a_n + a_{n+2} &= b_n, \\
&\text{(ix)} & c_n + b_{n-1} &= 2b_n, \\
&\text{(x)} & a_n + b_{n+2} &= 4a_{n+2}.
\end{aligned}
\tag{2}
$$

It is interesting to note here that the difference

$$a_n - a'_{n-1}, \tag{3}$$

gives the (slightly modified) Fibonacci sequence noted $\mathbf{F'_n}$

$$F'_n:\ 1, 0, 1, 1, 2, 3, 5, 8, 13, 21, 34, \dots, \tag{4}$$

in an unusual but interesting form: its "seeds" here are inverted with respect to the usual Fibonacci sequence. Also, the sum of any of its first members until a certain index gives a Fibonacci number, exactly, contrary to the usual Fibonacci sequence with seeds 0, 1 which always gives one unit less than a Fibonacci number. For example, in our case, for $\boldsymbol{n = 9}$, we get $\sum_1^9 \mathbf{F'_n} = \mathbf{34}$. (Note that the indexing is here shifted but the recurrence relation is still valid.) There is also another relation linking the sequences $\mathbf{a'_n}$ and $\mathbf{b_n}$. It writes

$$a'_n - b_{n-2} = 2F_{n-7}. \tag{5}$$

For $n = 7$, the sequences a' and b take the same value: $a'_7 - a_5 = 0$. Also, for $n = 8$, $a'_8 = 86$ and $b_6 = 84$ and their difference is 2. The case n = 9 is also interesting, see below. These relations will have applications in the following sections. Importantly, the sequences in Table 4 together with the one defined in Equ.(26), below, display several numbers highly relevant in this work, either directly as members in Table 4 (shown in bold) or as sums to be evaluated in the following sections. We have also discovered that the above sequences, including the one defined in Equ.(26), can all be shown to exhibit a striking *bilateral* symmetry and other symmetry properties, in the line of thought of those established for the Fibonacci sequence by Edge, [11]. These findings will be reported elsewhere, in a forthcoming publication.

4. The symmetries of the genetic code revealed

4.1. The multiplet structure

Let us consider, in this section, the first sequence a_n. Is is full of interesting numbers and sums. First, we have $a_4 = 8$, $a_5 = 15$ and their sum $a_4 + a_5 = 8 + 15 = 23$. This is the number of codons in Rumer's sets M_1 and M_2, respectively. Second, we have $a_6 + a_7 = 23 + 38 = a_8 = 61$. This is the pattern, "23 + 38", for 23 amino acids (the sextets counted two times) and 38 amino acids corresponding to 38 degenerate codons. This latter pattern will be mentioned, frequently, in this paper. In fact, the above relations will also let us to derive the detailed multiplet structure of the genetic code. As a matter of fact, consider the following sum, which will be used, occasionally, in this paper

$$\sum_1^k a_n = a_{k+2} - 1. \tag{6}$$

It is the analog of the one for the Fibonacci sequence. For $k = 5$, we have $6 + 1 + 7 + 8 + 15 = 38 - 1$. Grouping the first three terms, on the one hand, and the remaining two, on the other, we have by transferring the unit to the left

$$(6 + 1 + 7) + (8 + 1 + 15) = 14 + 24 = 38. \tag{7}$$

Using the sum mentioned above ($a_4 + a_5 = 8 + 15 = a_6 = 23$) and adding it to the preceding relation gives (by appropriately arranging the terms)

$$(15 + 14) + (8 + 24) = 29 + 32 = 61. \tag{8}$$

It appears that there are 15 amino acids and 14 degenerate codons in Rumer's set M_2 while there are 8 amino acids and 24 degenerate codons in Rumer's set M_1. Let us now go into the details by examining, first, the set M_2. The number 15 could be partitioned in two ways. The first, consists in using the above sum for k=3 to get: $6 + (1 + 7 + 1) = 6 + 9 = 15$. Using the second way, we can apply the useful A_0-function and its properties

(see below and Appendix B) to the number 15 (3×5): $A_0(15) = A_0(3) + A_0(5) = 6 + 9 = 15$ which gives the same result as above where we have used the *additivity* property. Finally, the number 6, a perfect number, could be written as the sum of its proper divisors: $6 = 1 + 2 + 3$, so that $15 = 1 + 2 + 3 + 9$. Here, we have 1 triplet, 2 singlets, 3 doublet parts of the 3 sextets, and 9 doublets. On the other hand, for the degeneracy part, 14, which writes $6 + 1 + 7$ (see above), we can, again, write 6 as the sum of its divisors, arrange the terms and obtain $14 = 3 + (1 + 1) + (2 + 7) = 3 + 2 + 9$. Here, we have 3 degenerate codons for the 3 doublet parts of the 3 sextets, 2 degenerate codons for the triplet and 9 degenerate codons for the 9 doublets. For the set M_1, things are simpler. The degeneracy part from Equ.(8) above writes $24 = (8 + 1) + 15 = 9 + 15$. As for the number of amino acids, 8, as a Fibonacci number, it could simply be written as $5 + 3$. This is the structure of the set M_1. The table below summarizes all these results for the two Rumer's sets which are thus completely described using the Fibonacci-like sequence a_n.

	multiplets	# amino acids	# degenerate codons	total
M_1	quartets	5	15	20
	quartet parts of the sextets	3	9	12
	total	8	24	32

	multiplets	# amino acids	# degenerate codons	total
	doublets	9	9	18
M_2	doublet parts of the sextets	3	3	6
	triplet	1	2	3
	singlets	2	0	2
	total	15	14	29

4.2. Hydrogen atom content and the symmetries

In this section, we examine the hydrogen atom content in each one of the symmetry cases summarized in section 2: Rumer's symmetry (section 2.1), the 3rd base symmetry (section 2.2) and the weak/strong, purine/pyrimidine and keto/amino symmetries (section 2.3). Before developing these topics, let us consider, first, the hydrogen atom content in the 61 amino acids side chains, as partitioned by the degeneracy. (Please note that when we say, here and below, "61 amino acids", we are of course taking into account the degeneracy of each one of the 20 amino acids.) From the table in the Appendix A, we have that the total number of hydrogen atoms in the 61 amino acids side chains is equal to 358. Let us note from the start that in this count, we take for the (singular) *imino acid proline*, as a special case, 5 hydrogen atoms in its side chain. We will return to this important point later, in section 5, with brand new results. A quick look at Table 4 of our Fibonacci-like sequences, reveals that the number of hydrogen atoms, mentioned above, is showing itself in multiple instances. First, ostensibly, as the ninth member of the sequence b_n ($b_9 = 358$). Second, from the relation (viii) in Equ.(2) which, we recall, is valid for any n, in particular for $n = 9$: $a_9 + a_{11} = 99 + 259 = 358 = b_9$. Third, from the recurence relation of the sequence b_n: $b_7 + b_8 = 137 + 221 = b_9 = 358$. Fourth, from the sum

$$\sum_1^9 a_n' = 358. \tag{9}$$

This last equation will be considered in detail below as it has a great importance concerning the computation of the degeneracy of the genetic code, in various formats. By isolating the last term a_9', we have

$$\sum_1^8 a_n' + a_9' = 219 + 139 = 358. \tag{10}$$

This relation is important and will play a prominent role is this section, and later (in section 6); it gives the number of hydrogen atoms in the amino acids side chains, distributed into two parts: 139 hydrogen atoms in 23 amino acids (17 amino acids with no "degeneracy" at the *first base* position and the six entities Ser^{IV-II}, Arg^{IV-II} and Leu^{IV-II}), on the one hand, and 219 hydrogen atoms in the 38 remaining amino acids corresponding to the 38 degenerate codons, on the other (see the introduction, below and also the Appendix A for the calculations). This is the pattern "23 + 38". Now, as we have $139 = 53 + 86 = 22 + 31 + 86$ from the recurrence relation of the sequence b_n, we can cast the relation above as follows

$$(219 + 22) + (31 + 86) = 241 + 117 = 358. \tag{11}$$

This is the hydrogen atom in the usual pattern "20 + 41" (20 amino acids and 41 degenerate codons). Note that 22 is the number of hydrogen atoms in serine, arginine and leucine, corresponding to one codon for each one of them (see the table in Appendix A). By restricting the sum in Equ.(10) as shown below, we have

$$\sum_1^7 a_n' + (a_8' + a_9') = 133 + 225 = 358. \tag{12}$$

This pattern corresponds to Rakocevic's *Cyclic Invariant Periodic System* (CIPS) classification of the 20 amino acids where there are 133 (225) hydrogen atoms in the amino acids side chains in the secondary superclass (primary superclass), see [12] and [13]. The above hydrogen atom pattern, $133 + 225$, is only one unit from another one which is twice relevant. By transferring the all first member of the sequence, $a_1' = 1$, from the sum to the other factor, we get

$$\sum_2^7 a_n' + (a_1' + a_8' + a_9') = 132 + 226 = 358. \tag{13}$$

First, this hydrogen atom pattern corresponds to 132 hydrogen atoms in the 3 sextets, on the one hand, and 226 hydrogen atoms in the remaining 17 amino acids, on the other (see below). Here, we see, the 3 sextets are set apart and this has, we think, a link with the subject of this section, see below. Second, this pattern describes also the partition of the 20 amino acids into 10 amino acids in the Class-I aminoacyl t-RNA synthetases (226 hydrogen atoms) and 10 amino acids in the Class-II aminoacyl t-RNA synthetases (132 hydrogen atoms), see [14]. Here, of course, the degeneracy is taken into account.

Now, in the following, we consider the hydrogen atom content for the three symmetry cases mentioned in sections 2.1, 2.2 and 2.3. As the case of section 2.3, the "ideal" symmetry classification scheme, [6], occupies an important place in this work, inasmuch as it has a tight relation with our "seeds" of Fibonacci-like series, we begin by considering it, first. As mentioned and promised in the introduction, it is here the right place to explain and justify the choice of the "seeds", or "initial conditions" of our Fibonacci-like sequences defined in section 3, more precisely those of the sequences b_n and c_n. Concerning the former, the "seeds" are 13 and 9 (see Table 4). These are, respectively the number of hydrogen atoms in serine and arginine $(10 + 3)$ and in leucine (9) and their sum, that is the recurrence relation, $b_1 + b_2 = 13 + 9 = b_3 = 22$, is the number of hydrogen atoms in the side chains of these three amino acids. As for the latter, the "seeds" are 30 and 5, respectively the number of atoms in the side chains of arginine and leucine $(17 + 13)$ and in the side chain of serine (5). Here also, as for hydrogen, $c_1 + c_2 = 17 + 13 = c_3 = 30$, is the number of atoms in the side chains of these three amino acids. We show, below in this paper, using all the resources offered by our Fibonacci-like series and their properties, that these 3 sextets, more precisely, their hydrogen and atoms numbers, as "seeds", will create the entire hydrogen atom, atom and even nucleon content of the whole set of amino acids, including the degeneracy, much like the creation of the 64 codons from the three sextets in the "ideal" symmetry scheme mentioned above, [6]. Let us, here in this section, begin with the hydrogen atom content. First, using the relation (v)

$c_n + 2b_{n-1} = b_{n+2}$ in Equ.(2), we can derive the hydrogen atom content in the two sets: the "leading" group and the "nonleading" group. As a matter of fact, for $n = 7$ (see Table 4), we have

$$190 + 2 \times 84 = 358. \tag{14}$$

It could be seen, from Table 3 and computed from the data in the table in Appendix A, that there are 190 and 168 hydrogen atoms in the"leading" group and in the "nonleading" group, respectively. Moreover, concerning the latter, we have that there are 84 hydrogen atoms in the amino acids the codons of which have the same first two bases UU, CC, AA and GG (in the four corners of Table 3) and 84 hydrogen atoms in the amino acids located in the four boxes in the center of the table which have as first two (different) bases UG, GU, AC and CA. The equation (14) above describes therefore, faithfully, this pattern. Now, we move further to describe accurately the hydrogen atom content involving the amino acids of the "core" comprising serine, arginine and leucine. To see this, we invoke the following two relations

$$5a_n + 2b_{n-1} = b_{n+2}, \tag{15}$$

$$3a_n + 4a_{n+1} = b_{n+2}. \tag{16}$$

It could be easily verified that they give the same result and both hold for any n. They can also be transformed into each other, using the relation (viii) in Equ.(2), $a_n + a_{n+2} = b_n$. Now, for $n = 7$, they give $190 + 168$ and $114 + 244$, respectively, with common value 358, the total number of hydrogen atoms in the 61 amino acids side chains. These are very interesting results for what follows. In the first case, as we have seen above, 190 is the number of hydrogen atoms in the "leading" group and 168 is the number of hydrogen atoms in the "nonleading" group. In the second case, 114 is the number of hydrogen atoms in the part of the "core" belonging to the "leading" group ($Ser^{IV/II}$, $Arg^{IV/II}$, Leu^{IV}) and 244 is the number of hydrogen atoms in the rest, comprising, in particular, the part of the "core" belonging to the "nonleading" group, that is, Leu^{II}. The authors write in their paper "the sextets as initial building blocks for creation of their new scheme of the genetic code generate by themselves the patterns of A+U rich/C+G rich, purine/pyrimidine, weak-strong and amino-keto symmetries." They also add that, in their approach, "the symmetries are a consequence of sextet's dynamics". Here, below, we can use our Fibonacci-like sequences to reveal the exact hydrogen atom content of the "core", constituted by the 3 sextets. As mentioned above, the "core" has two parts: the one which belongs to the "leading" group and the other belonging to the "nonleading" group. Let us consider the former with 114 hydrogen atoms. Using Euler's totient function φ and also so-called "reduced" totient function or Carmichael's function λ(n), see Appendix B, we have for the number 114 $\varphi(114) = 36$ and $\lambda(114) = 18$. Subtracting these from the number 114, we get $114 - 36 - 18 = 60$ and, rearranging, we get

$$114 = 60 + 36 + 18. \tag{17}$$

This is the correct content of the part of the "core" in the "leading" group: 60 hydrogen atoms (6×10) in arginine ($Arg^{IV/II}$), 36 hydrogen atoms (4×9) in leucine (Leu^{IV}) and 18 hydrogen atoms (6×3) in serine ($Ser^{IV/II}$). Let us, alternatively, add the above mentioned two functions to the number 114. We get

$$114 + 36 + 18 = (114 + 36) + 18 = 150 + 18 = 168. \tag{18}$$

This is the number of hydrogen atoms in the "nonleading" group where the isolated number 18 is now *re-interpreted* as the number of hydrogen atoms (2×9) in the "seed" of the "nonleading" group, that is, Leu^{II} (see above). We have thus established the exact hydrogen atom content of the "ideal" symmetry scheme of the genetic code where the sextets play a prominent role. Note, finally, that, as $\lambda(114) = 18$ has been used two times, one time as the number of hydrogen atoms in $\text{Ser}^{\text{IV/II}}$ and one time as as the number of hydrogen atoms in Leu^{II}, we can summarize all what has been said above by adding $\lambda(114)=18$ to Equ.(17) and write the exact hydrogen atom structure of the entire "core" $60 + (36 + 18) + 18 = 132$ constituted by $\text{Arg}^{\text{IV/II}}$, $(\text{Leu}^{\text{IV}} + \text{Leu}^{\text{II}})$ and $\text{Ser}^{\text{IV/II}}$, respectively. (The 18 codons of the "core" are undelined in Table 3.) Of course, subtracting the number 132 from the total sum 358, in Equ.(14) above, we are left with 226 which is the number of hydrogen atoms in the other part of the 17 amino acids outside the "core". We have seen above that the "seeds" of the sequences b_n and c_n are capable of creating the hydrogen atom structure of the "ideal" symmetry classification scheme. Now, what about the other sequences of Table 4 in this respect, that is the link to the "ideal" symmetry scheme? In fact, in turns out that they also house something interesting. Here, we consider only the sequence a_n but more will be said later in Section 4.3 (Equs.(34-35)), concerning the sequence g_n, defined below in Equ.(26). We have for the sequence a_n, using Equ.(6): $6 + 1 + 7 + 8 + 1 = 23$. First, from the Fibonacci relation $F_{2n} = F_{n+1}^2 - F_{n-1}^2$ with $n = 3$, we have $8 = 9 - 1$ or $8 + 1 = 9$. Second, it could be easily shown that the sequence F_n', in Equ.(4), is related to the Lucas sequence, $L_n = F_n' + F_{n+2}'$ so that, for $n = 5$, we have $7 = 2 + 5$. Finally, we call, exceptionally, the term $a_0 = -5$ which also obeys the recurrence relation $a_1 + a_0 = a_2$, that is $6 - 5 = 1$ or, equivalently, $6 = 5 + 1$. Putting together all these results, we end up with $(5 + 1) + 1 + 2 + 5 + 9 = 23$. The last four terms could be easily interpreted as 1 triplet, 2 singlets, 5 quartets and 9 doublets, that is the 17 amino acids outside the "core". As for the first two terms, in the parenthesis, they are just enough to describe the 5 entities $\text{Ser}^{\text{IV/II}}$, $\text{Arg}^{\text{IV/II}}$ and Leu^{IV} forming the part of the "core" belonging to the "leading" group, on the one hand, and 1 for Leu^{IV}, the part of the "core" belonging to the "nonleading" group, on the other.

As an interesting by-product of the results in this section, we have found, unexpectedly, a way to derive from the number of hydrogen atoms in the part of the "core" in the "leading" group, 114, and in the rest, 244, comprising the part of the "core" in the "nonleading" group (see above), and only from these, the very chemical structure of the building blocks of RNA; the four ribonucleotides *uridine monophosphate* (UMP), *cytidine monophosphate* (CMP), *adenosine monophosphate* (AMP) and *guanosine monophosphate* (GMP). As a matter of fact, using the functions A_0 and λ (see Appendix B), we have $A_0(114) = 38$, $A_0(244) = 88 = 61 + 1 + 18 + 4 + 4$ and $\lambda(114) = 18$ (see Appendix B where the details of the computations are given as examples). First, we have from these three quantities $[A_0(114) + \lambda(114)] + A_0(244) = 56 + 88 = 144$. This is the total number of atoms in the four ribonucleotides, 56 in the four nucleotides U (12 atoms), C (13 atoms), A (15 atoms) and G (16 atoms) and 88 in the four identical "blocks" each with 22 atoms (see [12], for the details of the calculation, which includes also a mathematical derivation of the number 22 above which is part of the "*condensation*" equation for the assembly of a ribonucleotide from the three units a nucleotide, a ribose and a phosphate group with the release of two water molecules, also derived). Now, as there are 30 codons in the "leading" group (two stop codons not counted) and 31 codons in the "nonleading" group (one stop codon also not counted), see Table 3, we can use this decomposition for the number 61 above and finally write the relations above in the form $(30 + 4) + (31 + 4) + (2 \times 18 + 1) + 38 = 34 + 35 + 37 + 38$. Note that the above decomposition of the number 61 could also be obtained another way, by using directly the properties of the sequence a_n, see Table 4. We have $a_8 = 61 = 23 + 38$, $a_7 = 38 = 23 + 15$ and $a_5 = 15 = 7 + 8$ so that by combining them, we get $61 = (23 + 7) + (23 + 8) = 30 + 31$. The above computed quantities 34, 35, 37 and 38 are respectively the number of atoms in the four ribonucleotides UMP, CMP, AMP and GMP (see [14]).

Now, we return to the symmetries and examine the second case, Rumer's symmetry (section 2.1). Let us reconsider Equ.(10) and write it in the form

$$\sum_1^7 a_n' + a_8' + a_9' = (133 + 53) + 2\times86 = 186 + 2\times86 = 358, \tag{19}$$

where we have used the recurrence relation of the sequence a_n', to write the number 139 as $86 + 53$ (see Table 4). Now, we have already mentioned in the examples following Equ.(5) that, for $n = 8$, one has $86 - 84 = 2$ or $86 = 84 + 2$. Inserting, this quantity in the above equation results in

$$186 + (84 + 88) = 358. \tag{20}$$

This is the hydrogen atom content in Rumer's division: 186 hydrogen atoms in M_2 and 172 hydrogen atoms in M_1 where, in this latter, we have the correct partition into 84 hydrogen atoms (4×21) in the 5 quartets and 88 hydrogen atoms (4×22) in the 3 quartets of the 3 sextets. To get the details concerning the number of hydrogen atoms in M_2, 186, we first isolate the sum of the first four numbers in the sum in Equ.(19), that is $1 + 6 + 7 + 13 = 27 = 3^3 = 3\times9$. This is equal to the number of hydrogen atoms in the triplet isoleucine (see below). We are left, in the sum, with the three terms 3×53. Writing one time the number 53 as $15 + 38$ from the relation (viii) in Equ.(2) with n=5 and two times as $22 + 31$ from the recurrence relation of the sequence b_n we obtain finally

$$2\times50 + 2\times22 + 27 + 7 + 8 = 186. \tag{21}$$

Here, $2\times50 = 2\times31 + 38 = 2\times(31 + 19)$ and $15 = 7 + 8$ from the recurrence relation of the sequence a_n. We have therefore, in the detail, the correct number of hydrogen atoms in M_2: $100 = 2\times50$ in the 9 doublets, $44 = 2\times22$ in the doublets of the 3 sextets, $27 = 3\times9$ in the triplet, 7 in the singlet methionine and 8 in the singlet tryptophane. Below, we consider the third and last case of symmetry. In section 2.2, we explained that the authors extracted an inherent basic symmetry linked to the third base by portioning the 64-codons set into four pair-wise sub-sets where each one of them contains only codons having the same third base. In this way, a one-to-one correspondence between one member of a doubly degenerate codon pair to the other member. Here also, for this symmetry, we could describe the hydrogen atom content, using our Fibonacci-like series. Take the relation (v) in Equ.(2), the one we already considered above in Equ.(14)

$$2\times84 + 190 = 358. \tag{22}$$

This relation, *as it is*, is the pattern shown in Table 2 for the gross third-base division UC/AG. Here, we note that this relation already describes, nicely, the equality of the number of hydrogen atoms in the columns 3rd base U and 3rd base C where the amino acids are exactly the same (see the penultimate row in Table 2). In fact, we can do better by invoking two more relations. First, the relation (x) in Equ.(2): $a_n + b_{n+2} = 4a_{n+2}$ which, for $n = 4$, gives $8 + 84 = 92$. Second, the relation $2b_n + b_{n+1} = c_{n+2}$, which also holds and gives, for $n = 5$, $2\times53 + 84 = 190$. Inserting the number $84 = 92 - 8$, from the relation just above, in the second one, results in $190 = 92 + 98$. Collecting these results in Equ.(22) above gives, finally

$$2\times84 + 92 + 98 = 358. \tag{23}$$

This last relation describes, therefore, completely, the hydrogen atom content pattern of Table 2. It is interesting that the 3rd base classification mentioned above can be supported by the following calculation. We know, from section 2.2, that the doubly degenerate codons (group-II) obey a fundamental symmetry, so they must play a basic role including,

we will show, in the hydrogen atom content too. As a matter of fact, we have, using the sequence a_n

$$\sum_1^9 a_n = 258. \tag{24}$$

Subtracting this sum from Equ.(22) above, which gives the total number of hydrogen atoms in the 61 amino acids, we get by arranging

$$100 + 258 = 358. \tag{25}$$

This is quite interesting as 100 and 258 are respectively the number of hydrogen atoms in the 9 doublets, on the one hand, and in the remaining 11 amino acids (5 quartets, 3 sextets, two singlets and 1 triplet), on the other, see below Equ.(21). In fact, this same relation could also be obtained, another way, from the relation mentioned in section 4.1, $a_9 + a_{11} = 99 + 259 = b_9 = 358$ and noting that the sum in Equ.(24) above is equal, in fact, to $259 - 1$ (recall $\sum_1^k a_n = a_{k+2} - 1$, with k=9), we get back to our result: $(1 + 99) + 258 = 100 + 258$. Note also that $2\times\varphi(258) = 2\times84$ and $358 - 2\times\varphi(258) = 190$ or $2\times84 + 190$ which is nothing but the hydrogen atoms pattern of the present classification (see Equ.(22) and Table 2). (The function φ is defined in **A**ppendix B and the factor 2 which has been introduced is for "doubly" degenerate codons.)

4.3. *The atom content and degeneracy*

In the course of writing this paper, we have discovered one more Fibonacci-like sequence, tailor-made for the description of the number of atoms in the 61 amino acids (see Equ.(29) below). It is defined as follows

$$g_n = -3F_{n-1} + 23F_{n-2}. \tag{26}$$

The first few terms are shown below

$$g_n: \ 23, -3, 20, 17, 37, 54, 91, 145, \mathbf{236}, 381, \dots \tag{27}$$

This sequence is related to the sequences a_n and b_n as follows

$$b_n + g_n = 6a_n, \tag{28}$$

which holds for any n. The case $n = 9$ is particularly interesting. As a matter of fact, we have

$$358 + 236 = 594, \tag{29}$$

and we see that it gives the total number of atoms in the 61 amino acids distributed into 358 hydrogen atoms (see above) and 236 atoms (C/N/O/S), see the table in Appendix A (180 carbon atoms and 56 N/O/S atoms). Now, from the relation

$$\sum_1^k g_n = g_{n+2} - g_2 = g_{n+2} - (-3) = g_{n+2} + 3, \tag{30}$$

which holds for any k and is the analog relation for the sum of the first k Fibonacci numbers. For $n = 7$, it gives $236 + 3$ or $236 = 239 - 3$ and, by inserting this latter in the above equation, we obtain

$$239 + (358 - 3) = 239 + 355 = 594. \tag{31}$$

We have here the number of atoms in 23 amino acids, 239 (the sextets with 35 atoms are counted two times) and 355 is the number of atoms in the 38 amino acids corresponding to the 38 degenerate codons (see the table in Appendix A). Let us, at this stage, remember the sequence c_n, especially its "seeds" 30 and 5 with sum $a_1 + a_2 = 35$. Let us also remember that these "seeds" were chosen, intentionally, as the sum of the number of atoms in Arginine and Leucine (30), on the one hand, and the number of atoms in Serine (5), on the other. They are therefore, together, exactly the right thing to add and subtract from Equ.(31) above to obtain

$$(239 - 35) + (355 + 35) = 204 + 390 = 594, \tag{32}$$

which is the correct partition of the number of atoms: 204 atoms in the 20 amino acids, on the one hand, and 390 atoms in the 41 amino acids corresponding to the 41 degenerate codons (see the table in Appendix A). Now, the use of the above sum in Equ(30), for $k = 8$, gives $\sum_1^8 g_n = 384$ which number appears also doubly significant, see below. By subtracting this latter number from the total sum, 594, and arranging, we have

$$210 + 384 = 594. \tag{33}$$

This partition of the number of atoms is interesting as 210 is equal to the number of atoms in the six amino acids (sextets) $\text{Ser}^{IV-II}, \text{Arg}^{IV-II}$ and Leu^{IV-II} (35×6) and 384 is the number of atoms in the remaining 17 amino acids, of course, taking into account the degeneracy. It is striking that the first two recurrence relations of the sequence g_n $23 - 3 = 20$ and $20 - 3 = 17$, together, lead to the relation

$$23 = 17 + (3 + 3), \tag{34}$$

which is in line with the above result for the atom numbers and also with the "ideal" symmetry scheme (as depicted below):

$$(3 + 3) \ \leftrightarrow \ (\text{Ser}^{IV}, \text{Arg}^{IV}, \text{Leu}^{IV}) + (\text{Ser}^{II}, \text{Arg}^{II}, \text{Leu}^{II}). \tag{35}$$

Finally, we could also derive the partition of the number of atoms for Rumer's sets M_1 and M_2. Consider, again, the equation above, $210 + 384 = 594$, more precisely, the number 384 which was calculated from Equ.(30) with $n = 8$. By partitioning this sum in two parts, for $n = 1..4$ and $n = 5..8$, we have $54 - (-3) = 54 + 3$, on the one hand, and 327, on the other. By gathering both terms in Equ.(33) and arranging, we obtain

$$(210 + 54) + (327 + 3) = 264 + 330 = 594. \tag{36}$$

This is the content of atoms 264 in M_1 and 330 in M_2, see the table in Appendix A. We can also go to the details for the multiplets. Considering, first, M_1, let us present the following (new) relation connecting the sequences b_n and c_n:

$$c_n + b_{n+2} = 4b_n, \tag{37}$$

which, for $n = 3$, gives $35 + 53 = 4×22 = 88$. Using the recurrence relation for b_n, we have $53 = 31 + 22$ and, by combining the above two relations, we get $35 + 31 + 22 = 4×22$ or $31 + 35 = 3×22 = 66$. Multiplying this latter equality by any number does not change it, in particular by 4, having in mind that the 8 quartets composing the set M_1 have each 4 codons, and we have $4×31 + 4×35 = 264$. This is exactly the number of atoms in M_1: $4×31$ in the 5 quartets and $4×35$ in the 3 quartet parts of the three sextets (see the table in Appendix A). The above equality, $31 + 35 = 66$, which was used as an intermediate of the calculation above, could also be exploited for the set M_2. As a matter of fact, consider Equ.(5), $a'_n - b_{n-2} = 2F_{n-7}$, for $n = 6$: $33 - 31 = 2$. Inserting this difference in the above equality gives $33 + 35 = 68$. Now, the following relation linking Fibonacci and Lucas numbers: $L_n + 3F_n = 2F_{n+2}$ is interesting as, for $n = 7$, it gives

$29 + 3\times13 = 2\times34 = 68$. If, moreover, we use the recurrence relation for the Lucas number 29 as $11 + 18$, we get $3\times13 + 11 + 18 = 68$. This matches perfectly the number of atoms in the triplet isoleucine (3×13), in the singlet methionine (11) and in the other singlet tryptophane (18), see the table in Appendix A. We showed above that there are 330 atoms in the set M_2. Subtracting the above number of atoms, 68, in the triplet and in the two singlets, we have 262 atoms left. To get the right partition of these, it suffices to take the sum of the three first members of the sequence c_n: $30 + 5 + 35 = 2\times35 = 70$ which appears to be the right number of atoms in the doublet parts of the three singlets. Adding and subtracting this latter from 262 gives 192 which is the number of atoms in the nine doublets, $2\times96 = 192$ (see the table in Appendix A). In summary, we have

$$M_1: 4\times31 + 4\times35 = 264, \tag{38}$$

$$M_2: 192 + 2\times35 + 3\times13 + 11 + 18 = 330,$$

which is the very precise partition. Finally, let us note that the number 384, mentioned above (see below Equ.(32)), has also another relevant interpretation. As a matter of fact, it is equal to the number of atoms in the 20 amino acids, this time including to their side chains their 20 identical "blocks" with 9 atoms each: $204 + 9\times20 = 384$.

4.4. Derivation of several nucleon number patterns

In this section, we use our Fibonacci-like series to derive several interesting patterns for the nucleon number (or integer molecular mass) content in the 61 amino acids. Before starting, let us make an important remark about the sequence c_n (see Table 4).. In fact, there is a simple relation between the sequences a_n and c_n, the latter is simply five times the former: $c_n = 5a_n$. One may wonder how the use of c_n would bring something significant as it is simply related to a_n? In fact, it does and we will show that below. First, let us consider the following sum

$$\sum_1^9 a_n + 2\sum_1^9 b_n + \sum_1^9 c_n = 3404. \tag{39}$$

It appears that this number, 3404, is the number of nucleons, or integer molecular mass, contained in the 61 amino acids (see the table in Appendix A). This is nice but we could do more. Consider the "seeds" of the sequence c_n, 30 and 5 with sum 35, the number of atoms in the 3 sextets serine (5), arginine (17) and leucine (13). (It is not difficult to show, using the Zekendorf representation[1] of the number 30 ($= 21 + 8 + 1$) in terms of Fibonacci numbers and the fact that $21 = 13 + 8$, that the sum of the "seeds" takes the form $13 + 17 + 5 = 35$, i.e., the correct atom numbers in the three sextets, mentioned above.) Now, by isolating the sum of the above two "seeds" of c_n from the third sum in Equ.(39) and include it in the two other sums, we get

$$2149 + 1255 = 3404. \tag{40}$$

Here also, we have another interesting result: there are 1255 nucleons in the 20 amino acids (see the table in Appendix A) and 2149 nucleons in the 41 amino acids corresponding to 41 degenerate codons. Let us now exploit the relation between the two sequences a_n and c_n ($c_n = 5a_n$), mentioned above, and write the sum in Equ(39) as follows

---

[1] The Zeckendorf theorem states that every positive integer can be represented uniquely as the sum of one or more non-consecutive Fibonacci numbers.

$$\left[4\sum_1^9 a_n + \sum_1^9 b_n \right] + \left[2\sum_1^9 a_n + \sum_1^9 b_n \right] = 1960 + 1444 = 3404. \tag{41}$$

Now, recall the sum $\sum_1^k a_n = a_{k+2} - 1$, mentioned in Equ.(6) of section 4.1. In our present case for its use in Equ.(41), we have $\sum_1^9 a_n = 259 - 1$ for $k = 9$. By considering this latter relation in *only one* such sum in the first bracket of the above equation and transfer the unit, -1, to the second bracket, we obtain

$$1961 + 1443 = 3404. \tag{42}$$

One recognizes here the nucleon number in the pattern "38 + 23" (see above and Appendix A): 1443 nucleons in the 23 amino acids side chains and 1961 nucleons in the 38 amino acids side chains corresponding to the 38 degenerate codons (see above). As another, also interesting finding, we can, from the above relations, make contact with the "ideal" symmetry mentioned above in section 4.2, at the level of the nucleon numbers. To do this, let us, first, remark that the number 114 appears *two times*, one time as the number of hydrogen atoms in the "core" of the "ideal" symmetry scheme (see above) and one time as the number of nucleons in Leu^{II} (2×57), see the table in Appendix A). This will prove significant in the following. Consider now the sum

$$\sum_1^9 a_n + 2\sum_1^9 b_n = 2114. \tag{43}$$

The number 2114, by itself, is not very interesting but its φ-function is. As a matter of fact, we have $\varphi(2114) = 900$ and, adding to it two times the number 114 gives $900 + 2{\times}114 = 1128$. This is the number of nucleons in the "core": $31{\times}6 + 100{\times}6 + 57{\times}6 = 1128$. Isolating one time the number 114 in the sum, that is $(900 + 114) + 114 = 1014 + 114$ gives us the partition of the nucleon numbers between the two parts of the "core", $31{\times}6 + 100{\times}6 + 57{\times}4 = 1014$ in the "leading" group, on the one hand, and $57{\times}2 = 114$ in the "nonleading" group, on the other:

$$1128 = 1014 + 114. \tag{44}$$

In the following, we can also derive three more results by "watering three plants with one hose", so to speak. As a matter of fact, consider again the sum in Equ.(39) and cut it as follows

$$\left[\sum_1^9 a_n + \sum_1^9 b_n + \sum_1^7 c_n \right] + \left[\sum_1^9 b_n + \sum_8^9 c_n \right] = 1676 + 1728 = 3404. \tag{45}$$

Here, we have the nucleon number pattern of the 3rd-base classification of section 2.2: 1728 nucleons in the U/C third-base division and 1676 nucleons in the A/G third-base division (see Table 2, last row). Now, by borrowing from the first bracket above the sum of the first three members of the sequence c_n: $30 + 5 + 35 = 2{\times}35 = 70$, the one we used earlier (see above Equ.(38)), to the benefit of the second bracket, we get[2]

$$1606 + 1798 = 3404. \tag{46}$$

Here, we recognize the number of nucleons in the "leading"group/"nonleading" group, 1606 and 1798, respectively. Finally, we could also establish the nucleon number pat-

---

[2] As an example of the computations, here is the one for the « nonleading » group which goes as follows: $31{\times}6 + 57{\times}4 + 100{\times}6 + 15{\times}4 + 59{\times}2 + 73{\times}2 + 107{\times}2 + 57{\times}3 + 75 = 1798$, see Table 3 and the table in Appendix B.

tern corresponding to Rumer's division. Consider again Equ.(39) by partitioning it as follows

$$[\sum_1^8 a_n + 2 \sum_1^8 b_n + \sum_1^8 c_n] + (a_9 + 2b_9 + c_9) = 2094 + 1310 = 3404. \qquad (47)$$

It suffices now, analogously to what we did in Equ. **(40)** above, to subtract, *one time*, the sum of the "seeds" of the sequence b_n in the bracket**,** that is, **13 + 9 = 22**, and add it to the three terms in the parenthesis to obtain

$$2072 + 1332 = 3404. \qquad (48)$$

We have, as promised above, 1332 nucleons in M_1 and 2072 nucleons in M_2 (see the table in Appendix A).

5. On proline's singularity and a derivation of the shCherbak-Makukov "activation" key

In this section, we use our Fibonacci-like sequences to shed light, by giving concrete results, on a question relative to the special amino (more exactly imino) acid, proline, which is an exception among the set of 20 amino acids. As a matter of fact, it is the only amino acid where its side chain is connected to its block twice. shCherbak, [7], in order to "standardize" the common block of the amino acids, with 74 nucleons, proposed an *imaginary* "borrowing" of one nucleon (in fact one hydrogen atom) from the side chain of proline, which has only 73 nucleons in its block, to the benefit of its block, to reach 74, as the 19 other amino acids. In his next work with Makukov, [8], the above "borrowing", or transfer of one nucleon, has been termed the "*activation key*". Activating the key, *i.e.*, standardizing, leads to an innumerable number of remarkable and beautiful arithmetical patterns. These authors say in their paper, [8], "*Applied systematically without exceptions, the artificial transfer in proline enables holistic and arithmetically precise order in the code*". Here, in this section, we establish,, not only a mathematical version of the "activation key", itself, but also its action on the (new) total hydrogen atom content, with simple possible extension to the atom and nucleon content. Let us begin by examining the action of the "activation key"). Consider, again, the sequence a'_n and the following sum

$$\sum_1^k a'_n = a'_{k+2} - 6. \qquad (49)$$

It could be easily shown and verified that the above relation holds for any k. For k=9, it gives $358 = 364 - 6$. As established and mentioned many times previously, 358 is the number of hydrogen atoms in the 61 amino acids side chains, where the special amino acid proline has 5 hydrogen atoms in its side chain. If, instead, one considers that proline's side chain has now 6 hydrogen atoms, at the cost of its block, *i.e.*, no standardization made, or the "activation key" *off* (see below), and taking into account the number of its coding codons, which is 4, then we have now $362 = 358 + 4$ hydrogen atoms in the 61 amino acids side chains. Let us reconsider Equ.(10)) for the hydrogen atoms partition between 38 amino acids corresponding to the 38 degenerate codons (219) and 23 amino acids (139) but, now, using the above relation ($358 = 364 - 6$):

$$\sum_1^8 a'_n + a'_9 = 219 + 139 = 364 - 6. \qquad (50)$$

To get a correct partition, let us consider the *perfect number* 6 which is, as such, equal to the sum of its proper divisors: **6 = 1 + 2 + 3**. These are just the right numbers we need. Inserting these in the above equation by selecting the odd divisors **1** and 3 and shifting

them to the left while leaving the even one, 2, to the right and finally arranging properly, we get

$$\sum_1^8 a'_n + a'_9 = (219 + 3) + (139 + 1) = 364 - 2 = 362, \tag{51}$$

so that we have something quite correct: one more hydrogen atom in the 23 amino acids part (proline, a quartet, has now one more hydrogen atom in its side chain) and 3 more hydrogen atoms for its 3 degenerate codons. (Note that, the action of the "activation key" could easily be extended to the total number of atoms and the total number of nucleons, as it applies only to the hydrogen atom number content.) Taking a look at the 6th term in the sequence c_n, $115 = 40 + 75$, we have that it appears to be equal to the number of nucleons in proline *side chain and block*, see below about this latter sum. This number, 115, is *invariant* whether we make shCherbak's "borrowing" of one nucleon or not. To get more insight, we consider another invariant number, the total number of hydrogen atoms in the 61 amino acids, *including* the blocks (with 4 hydrogen atoms in each), that is $358 + 244 = 362 + 240 = 602$. Without borrowing one nucleon from the side chain of proline in favour of its block there are 362 hydrogen atoms in the 61 amino acids and 240 hydrogen atoms, $57{\times}4 + 4{\times}3 = 240$, in the 61 blocks. Applying the "borrowing", there are 358 hydrogen atoms in the 61 amino acids side chains and $244(= 61{\times}4)$ hydrogen atoms in the 61 blocks. Note, in passing, the following nice relations seemingly linking the two views: $\varphi(240) + \varphi(362) = 244$ and $(240 + 362) - [\varphi(240) + \varphi(362)] = 358$. Now, let us examine the former point,, the derivation of the "activation key". Considering the above mentioned invariant numbers, $115\ (= 5{\times}23))$ and $602\ (= 2{\times}7{\times}43)$, we have, using their A_0 function (see Appendix B):

$$115 - A_0(115) = 115 - 42 = 73, \tag{52}$$

$$115 - A_0(602) = 115 - 74 = 41. \tag{53}$$

From which we deduce that $115 = 42 + 73 = 41 + 74$ which are seen to describe, fully and precisely, the two views: $42 + 73$ ("activation key" *off*) and $41 + 74$ ("activation key" *on*). From $\sigma(41) = 42 = 41 + 1$, where σ is the sum of the divisors, we can also write $115 = (41 + 1) + 73 = 41 + (73 + 1)$. Also, from $\varphi(41) = 40 = 41 - 1$, we can also make contact with the sequence c_n through the relation $c_6 = 115 = 40 + 75$, mentioned above: $41 + (75 - 1) = 41 + 74$. Moreover, we can, alternatively, exploit the number 75, itself. As a matter of fact, calling Legendre's sum of three squares theorem[3], we have that the number 75 do not satisfy the theorem and can therefore be written as the sum of the following three squares: $1^2 + 5^2 + 7^2$ or $1 + (25 + 49) = 1 + 74$. This latter form gives us again $(40 + 1) + 74 = 41 + 74$. Finally, using $\varphi(41) = 40 = 41 - 1$ and the decomposition of the number 75 as the sum of three squares, mentioned above, we can write, by allocating the two units in two ways: $41 - 1 + 1 + 74 = 41 + 74 = 42 + 73$. This is, again, what we found above from Equs.(52-53).

6. A striking imprint in the "seeds"

We invite the reader, here, before starting this section, to remember what has been said about the three sextets in section 4.2 when using the "seeds" 6 and 1 of our fifth sequence $\mathbf{g_n}$. In the "ideal" symmetry classification scheme, mentioned in section 2.3, the authors explain that, in their approach, the symmetries are a consequence of sextet's dynamics and the whole set of amino acids is created starting from these three sextets

---

[3] This theorem states that a natural number n can be represented as a sum of three squares if and only if it is not of the form $4^a(8b + 7)$ for a and b two positive integers. It could be easily verified that the number 75 cannot be written in this form so it can be represented as the sum of three squares.

where, serine, plays a prominent role. In our own approach, relying on the use of Fibonacci-like series, on the other hand, we have chosen, as already mentioned, for two of them, b_n and c_n, the hydrogen atom numbers and atom numbers of the three sextets (see section 4.2) as "seeds". It appears, while the writing of this paper goes to its end, that the "seeds" of, in fact, *all* the Fibonacci-like sequences used in this paper, and only these, by themselves, can "create", strikingly and remarkably, the main hydrogen number patterns derived in this paper. As a matter of fact, the *sum and product* of the "seeds" of the sequence b_n, *alone*, gives

$$b_1 \times b_2 + (b_1 + b_2) = 117 + 22 = 139. \tag{54}$$

One recognize here the number of hydrogen atoms in 20 amino acids, 117, augmented by the number of hydrogen atoms in the three sextets, 22. The total, 139, corresponds to 23 amino acids (the sextets counted two times). Now, let us compute the following expression, using the *sum and product* of the "seeds" of the sequence $\mathbf{c_n}$ and *only* the sum of the "seeds" of the other three remaining sequences a_n, a'_n, and g_n (the latter defined in Equ.(26). We have

$$c_1 \times c_2 + (c_1 + c_2) + (a_1 + a_2 + a'_1 + a'_2) + (g_1 + g_2) =$$

$$= 150 + 35 + 14 + 20 = 219. \tag{55}$$

Here, we have the number of hydrogen atoms in the 38 amino acids corresponding to the 38 degenerate codons. Equs.(54-55), together, constitute the "23 + 38" hydrogen atom pattern established in section 4.1). Now, taking the sum of the above equations and borrowing the number 22 from Equ.(54) to the benefit of Equ.(55) gives $117 + 241 = 358$ which corresponds to the other pattern "20 + 41" (20 amino acids and 41 degenerate codons), see Equ.(11) in section 4.1. Next, we arrange Equs.(54) and (**55**) as follows

$$(150 + 22) + (117 + 35 + 14 + 20) = 172 + 186 = 358. \tag{56}$$

Here, we have again the hydrogen atom content in Rumer's division: 172 hydrogen atoms in M_1 and 186 hydrogen atoms in M_2, see section 4.2 Equ.(20). To get the other patterns, we call, exceptionally, the Fibonacci $(0, 1, 1, 2, 3, 5, \dots)$ and Lucas $(2, 1, 3, 4, 7, 11, \dots)$ series, which, as well known, are linked by the relation $F_n + L_{n+2} = F_{n+4}$. For n=5, we have $5 + 29 = 34$ so that we can take this latter as the term $14 + 20 = 34$ in the above equations, arrange and get

$$(150 + 35 + 5) + (22 + 117 + 29) = 190 + 168 = 358. \tag{57}$$

This is the hydrogen atom pattern for (i) the 3rd base classification of section (4.2), Equ.(14), and (ii) the "ideal" symmetry classification scheme in the same section, Equ. (22). Finally, we consider Zekendorf's theorem (see footnote 1) and apply it to the number 117 giving $89 + 21 + 5 + 2$. Writing 89, a Fibonacci number, as $55 + 34$, we can rearrange the content of the second parenthesis in Equ.(57) above as $55 + 29 = 84$ and $34 + 21 + 22 + 5 + 2 = 84$, so that $168 = 2 \times 84$, which describes again the pattern $190 + 2 \times 84 = 358$. The fact of having used, here, the Fibonacci and Lucas sequences is all the more interesting in that it can, also, give us another remarkable result. As a matter of fact, adding the two "seeds" of the Fibonacci and Lucas sequences, 0 and 1 and 2 and 1, respectively, to the above sum of Equs.(54) and (55) and arranging, we obtain

$$(139 + 1) + (219 + 2 + 1) = 140 + 222 = 362. \tag{58}$$

This is, exactly, the hydrogen atom pattern found in section (5), devoted to the special imino acid proline and the shCherbak-Makukov "activation" key, when this latter is "*off*", see Equ.(51) and below it**,** in section 5.

Conclusion

In this work, we have strayed a little off the beaten paths in the genetic code mathematical research. Starting with a handful of Fibonacci-like sequences, we have derived not only the degeneracy structure of the genetic code but also the hydrogen atom content, the atom number content and also the integer molecular mass (nucleon) content of the set of 20 amino acids as strutured in the 64-codon table. What made this possible is a judicious choice of the intial conditions, or "seeds", of the above mentioned sequences. For two of these, they are chosen as the hydrogen atom numbers and atom numbers of the three enigmatic and intriguing amino acids serine, arginine and leucine. Our results, using these sequences, led us to reveal, as mentioned above, the elemental content of the 61 amino acids set as structured by various well-known symmetries, as Rumer's symmetry, the "ideal" symmetry classification scheme and the basic symmetry associated to order-2 degeneracy. Moreover, as a by-product of our mathematical formalism, we derived the atomic (elemental) content of the building-blocks of RNA, the four ribonucleotides UMP, CMP, AMP and GMP. Also, still using the above mathematics, we bring, for the first time, an additional brick to shCherbak's theory concerning the role of the special imino acid proline whose virtual "double" struture renders possible, via the use of the "activation key", a large number of remarkable and beautiful arithmetical patterns. Let us stress, as a last word, that the content of this paper is, we believe, inovative. To the best of our knowledge, we have never seen such a kind of quantitative derivation of the chemical characteristics of the genetic code as structured by the degeneracy and also by several well known symmetries. Our main findings, as the total hydrogen atom content, the total atom content, the total molecular mass content of the 20 amino acids, including the degeneracy, as well as other relevant quantities related to the symmetries of the genetic code are found directly, either as ostensible members of the Fibonacci-like sequences or obtained from the summation properties of these latter. Other slighly more sophisticated quantities, also having something to do with the symmetries, are obtained with the help of some well known arithmetic functions.

Appendix A

In the table of this appendix, we give the detailed elemental composition of the side chains of the 20 amino acids. H stands for hydrogen, C for carbon, N for nitrogen, O for oxygen and S for sulfur. The calculated values of some important quantities, taking into account the degeneracies, are indicated in the last five rows; they are useful to know when reading the main text (*those shown in bold, and others, are all mathematically derived in this paper*). In the table, the first column, M, gives the number of codons which code for an amino acid (4 for a quartet, 6 for a sextet, 2 for a doublet, 3 for a triplet and 1 for a singlet). In column 6 we provide the number of atoms in the side chains and the number of nucleons (protons and neutrons), which is also the integer molecular mass of an amino acid, is displayed in column 7. Below the table, we offer hints for computing some of them. The table is in the "standardized" form, that is, proline has 5 hydrogen atoms in its side chain and all the twenty amino acids, including proline, have 74 nucleons in each of their blocks, see section 5. The general chemical (linear) formula of an amino acid is

$$R - \mathbf{C}H(NH2) - COOH,$$

where R is the radical, also called the side chain, and the rest of the molecule constitutes what we call here the block. The carbon C, linked to R, is called α-carbon. In the special case of proline, its side chain from the α-carbon connects to the nitrogen N forming a

pyrrolidine loop. (It is the side chain which gives an amino acid its specific functional properties.). To calculate, for example, the nucleon numbers, or the integer molecular mass of an amino acid, the molecular mass of the chemical elements are those of the most abundant isotopes: hydrogen (1), carbon (12), nitrogen (14), oxygen (16) and sulfur (32). From the formula above, one computes easily the integer molecular mass of the block: $2 \times 12 + 1 \times 14 + 2 \times 16 + 4 \times 1 = 74$. In the (*unique*) case of proline, as mentioned above, there is one less hydrogen atom in the block and the nucleon number is $73 = 74 - 1$; this is the non-standardized form ("activation key" *off*), (see section 5).

M	amino acid	# H	# C	# N/O/S	# atoms	# nucleons
4	Proline (P)	5	3	0	8	41
	Alanine (A)	3	1	0	4	15
	Threonine (T)	5	2	0/1/0	8	45
	Valine (V)	7	3	0	10	43
	Glycine (G)	1	0	0	1	1
6	Serine (S)	3	1	0/1/0	5	31
	Leucine (L)	9	4	0	13	57
	Arginine (R)	10	4	3/0/0	17	100
2	Phenylalanine (F)	7	7	0	14	91
	Tyrosine (Y)	7	7	0/1/0	15	107
	Cysteine (C)	3	1	0/0/1	5	47
	Histidine (H)	5	4	2/0/0	11	81
	Glutamine (Q)	6	3	1/1/0	11	72
	Asparagine (N)	4	2	1/1/0	8	58
	Lysine (K)	10	4	1/0/0	15	72
	Aspartic Acid (D)	3	2	0/2/0	7	59
	Glutamic Acid (E)	5	3	0/2/0	10	73
3	Isoleucine (I)	9	4	0	13	57
1	Methionine (M)	7	3	0/0/1	11	75
1	Tryptophane (W)	8	9	1/0/0	18	130
	Total (20)	**117**	67	20	**204**	**1255**
	Total (23)	**139**	76	24	**239**	**1443**
	Total (38)	**219**	104	32	**355**	**1961**
	Total (61)	**358**	180	56	**594**	**3404**
	M_1/M_2	172/186			264/330	1332/2072

The get the results in the second of the last five rows from the first one, it suffices to count the values of the sextets two times. For the rest, to ease the calculations, one can use the following pre-calculated sums for the hydrogen atom content: 5 quartets 21, 3 sextets 22, 9 doublets 50, 1 triplet 9, 2 singlets $15 = 7 + 8$. For the atom number: 5 quartets 31, 3 sextets 35, 9 doublets 96, 1 triplet 13, 2 singlets $29 = 11 + 18$. For the nucleon numbers: 5 quartets 145, 3 sextets 188, 9 doublets 660, 1 triplet 57, 2 singlets $205 = 75 + 130$.

In the calculations, the reader needs also to know what we mean by degeneracy. This latter is defined as the number of codons coding for an amino acid *minus one*. Therefore, for a quartet, the degeneracy is $3 = 4 - 1$, for a doublet it is $1 = 2 - 1$, for a triplet it is $2 = 3 - 1$ and for singlet it is $0 = 1 - 1$. For the special case of the sextets, there are two possibilities, related to the two patterns mentioned several times in this paper: "20 + 41 = 61" and "23 + 38 = 61". In the first case, the degeneracy is $3 + 2 = 5$ (3 for the quartet part and 2 for the doublet part whose two codons are considered both degenerate). In the second case, the quartet part and the doublet part of each sextet are considered as *separate* entities (ex. Ser^{IV} and Ser^{II}) so the degeneracy is equal to $3 + 1 = 4$, 3 for the quartet part and 1 for the doublet part which is here considered as a doublet. In this way, we have for the number of amino acids and total number of coding codons $20 = 5 + 3 + 9 + 1 + 2$, $41 = 5 \times 3 + 3 \times 5 + 9 \times 1 + 1 \times 2$, in the first case, and $23 = 5 + (3 + 3) + 9 + 1 + 2$, $38 = 5 \times 3 + 3 \times (3 + 1) + 9 \times 1 + 1 \times 2$ in the second one. With these definitions, it is not difficult to do the rest of the computations. Let us give some few

examples from the table above for the number of hydrogen atoms for the pattern"23 + 38: $139 = 21 + 22 \times 2 + 50 + 9 + 7 + 8$, $219 = 21 \times 3 + 22 \times 4 + 50 \times 1 + 9 \times 2$, $358 = 21 \times 4 + 22 \times 6 + 50 \times 2 + 9 \times 3 + 7 + 8$.

Appendix B

In this appendix, we mention some few other additional mathematical elements used in this paper. First, Euler's totient function for an inger n, $\varphi(n)$, which is extensively used in many scientific areas as cryptography and graph theory, as a few examples. It counts the number of positive integers less than or equal to n which are relatively prime to n (also called coprimes). For example 24 has 8 coprimes (1, 5, 7, 11, 13, 17, 19, 23): $\varphi(24) = 8$. Second, Carmichael λ-function, also called *reduced* totient function, which is, in fact, used only one time in section 4.2 where it appears to be very useful. It is defined as the smallest positive divisor of Euler's totient function that satisfies Euler's Theorem. For example $\lambda(24) = 2$. (The reader could easily find good online calculators for these functions to check.) Next, the A_0 function which has been used, successfully, by us many times in our recent works on the genetic code, see for example [13]. It is define by

$$A_0(n) \coloneqq a_0(n) + SPI(n) + \Omega(n),$$

where $a_0(n)$ is the sum of the prime factors, $(p_1 \times n_1 + p_2 \times n_2 + \cdots + p_k \times n_k)$ of the integer n (written as $p_1^{n_1} \times p_2^{n_2} \ldots \times p_k^{n_k}$ (by the *Fundamental Theorem of Arithmetic*), including the multiplicities, $SPI(n)$ is the Sum of the Prime Indices $PI(p_1) \times n_1 + PI(p_2) \times n_2 + \cdots + PI(p_k) \times n_k$ where PI(2)=1, PI(3)=2, PI(5)=3 and so on, also including the multiplicities and $\Omega(n)$, so-called Big Omega function, is the number of the prime factors $n_1 + n_2 + \cdots + n_k$. Consider, as an example, the number 192 whose prime fctorization is $2^6 \times 3^1$. We have

$$A_0(192) = a_0(2^6 \times 3^1) + SPI(2^6 \times 3^1) + \Omega(2^6 \times 3^1)$$

$$= (6 \times 2 + 1 \times 3) + (6 \times 1 + 1 \times 2) + (6 + 1) = 30.$$

The function A_0 enjoys also the useful additivity ("logarithmic") property $A_0(n \times m \times p \times \ldots) = A_0(n) + A_0(m) + A_0(p) + \cdots$. Let us also give some few other illustration examples, taken from section 4.2, concerning the computation of $A_0(114)$ and $A_0(244)$. For the first, we have $114 = 2^1 \times 3^1 \times 19^1$ so that $A_0(114) = (2 + 3 + 19) + (1 + 2 + 8) + 3 = 38$. For the second, we have $244 = 2^2 \times 61^1$. To get the result established in the end of section 4.2, it makes sense to use the additivity property mentioned above: $A_0(244) = A_0(2^1) + A_0(2^1) + A_0(61^1) = 4 + 4 + (61 + 18 + 1) = 88$. This form, which sets apart the two factors 4 proved useful to reveal the structure of the four ribonucleotides (in section 4.2).

References

1. Nirenberg, M.; Leder, P.; Bernfield, M.; Brimacombe, R.; Trupin, J.; Rottman, F.; O'Neal, C.N.A. Codewords and Protein Synthesis, VII. On the General Nature of the RNA Code. *Proc Nat. Acad Sci USA* 1965, *53(5)*, 1161-1168.
2. Négadi, T. The genetic code degeneracy and the amino acids chemical composition are connected. *NeuroQuantology* 2009,7(1), 181-187
3. Inouye, M.; Takino, R.; Ishida, Y.; Inouye, K. Evolution of the genetic code; Evidence from codon use disparity in *Escherichia coli. Proc Natl Acad Sci USA* 2020, 117(46), 28572-28575.
4. Zwick, A.; Regier, J.C.; Zwickl, D. Resolving Discrepancy between Nucleotides and Amino Acids in Deep-Level Arthropod Phylogenomics: Differentiating Serine Codons in 21-Amino-Acid Models. *PLoS ONE* 2012, 7(11), e 47450
5. Sun, X.; Yang, Q.; Xia, X. An improved implementation of effective number of codons. (Nc) *Mol Biol Evol* 2013, 30(1), 191-196.
6. Rosandić, M.; Parr, V. Codons sextets with leading role of serine create "ideal" symmetry classification scheme of the genetic code. *Gene* 2014, 543(1), 45-52.

7.   shCherbak, V. The Arithmetical origin of the genetic code. In The codes of Life: The rules of Macroevolution. Barbieri, M.; Ed. New York: Springer Publishers, 2008, pp. 153-185.

8.   shCherbak, V.; Makukov, M. The "wow! Signal" of the terrestrial genetic code. _Icarus_ 2013, 224, 228-242.

9.   Rumer, Y. About systematization of the genetic code. _Dok. Akad Nauk SSSR_ 1966, 167, 1393-1394.

10.   Findley, G.I.; Findley, A.M.; McGlynn, S.P.. Symmetry characteristics of the genetic code. _Proceeding of the National Academy of Sciences of the United States of America_, 79, 7061-7065.

11.   Edge, M. Symmetry in Fibonacci numbers. _Symmetry: Culture and Science_ 2009, 20(1-4), 393-408.

12.   Rakocevic, M.M. Genetic Code: The unity of the stereochemical determinism and pure chance. 2009, arXiv:0904.1161v1 [q-bio.OT]

13.   Négadi, T. A Taylor-made arithmetic model of the genetic code and applications. _Symmetry: Culture and Science_ 2009, 20(1-4), 51-76.

14.   Négadi, T. The genetic code invariance: when Euler and Fibonacci meet. _Symmetry: Culture and Science_ 2014, 30(2), 261-2