

FIBONACCI NUMBERS IN RNA IMPLY THE ONE NUMBER MODEL OF THE GENETIC CODE

Tidjani Négadi

Physics Department, Faculty of Science, University of Oran, 31100, Oran, Algeria; E-mail: tnegadi@gmail.com

In this short paper, we present an interesting and welcome connection between two different approaches to the mathematical structure of the standard genetic code, we have considered in the last years. The first one relies on the use of the unique number $23!$ and the second is based on the atomic composition of the four ribonucleotides UMP, CMP, AMP and GMP, the building-blocks of RNA, where several Fibonacci numbers are seen to occur.

In 2007 (Négadi, 2007), we have presented, in this journal and for the first time, a mathematical model of the genetic code based on the unique number

$$23! = 25852016738884976640000 \quad (1)$$

This latter number has twenty-three digits, in the decimal place-value representation, eighteen of them comprising the digits from 1 to 9 (with repetitions) and five zeros. Also, this number has the following (unique) prime-factorization, from the Fundamental Theorem of Arithmetic

$$23! = 2^{19} \times 3^9 \times 5^4 \times 7^3 \times 11^2 \times 13 \times 17 \times 19 \times 23 \quad (2)$$

In Eq.(2), one counts forty-one factors (the sum of the exponents). It is well known that the 64 triplet-codons of the Standard Genetic Code are organized into five *multiplets* of amino acids: 5 *quartets* (total 20 codons), Proline (P), Alanine (A), Threonine (T), Valine (V), Glycine (G); 3 *sextets* (total 18 codons) Serine (S), Leucine (L), Arginine (R); 9 *doublets* (total 18 codons) Phenylalanine (F), Tyrosine (Y), Cysteine (C), Histidine (H), Glutamine (Q), Asparagine (N), Lysine (K), Glutamic Acid (E), Aspartic Acid (D); 1 *triplet* (3 codons) Isoleucine and finally 2 *singlets* (1 codon each) Methionine (M) and Tryptophane (W). There are 64 codons in all and 20 amino acids

coded so 41 codons are said “degenerate”. The following sorting of the digits in Eq.(1) fits exactly the above multiplet structure of the standard genetic code (Négadi, 2007)

$$\begin{aligned}
 &5 \text{ “quartets”}: \{3, 5, 5, 7, 7\}; \\
 &3 \text{ “sextets”}: \{1, 2, 9\}; \\
 &9 \text{ “doublets”}: \{4, 4, 6, 6, 6, 8, 8, 8, 8\}; \\
 &1 \text{ “triplet”}: \{2\}; \\
 &2 \text{ “singlets”}: \{0, 0\}; \\
 &3 \text{ “stops”}: \{0, 0, 0\}
 \end{aligned} \tag{3}$$

Our computations, as usual (see Négadi, 2007, 2008, 2009a, b), rely on the use of some elementary arithmetic functions, among them, $a_0(n)$ the sum of the prime factors of a number n from its prime factorization via the Fundamental Theorem of Arithmetic (including multiplicities), $SPI(n)$ the sum of their prime indices and $\Omega(n)$ the number of the prime factors (the Big Omega Function). We also define A_0 as the sum of these three functions: $A_0(n)=a_0(n)+SPI(n)+\Omega(n)$. In Eq.(3), the number of digits represent the *number* of amino acids in a given “multiplet”. The degeneracy, *i.e.*, the number of degenerate codons, in a given multiplet is then computed as the sum of the a_0 -functions of the factorized digits in a multiplet (without counting multiplicities). Using these latter functions, we have derived many interesting and relevant results concerning the mathematical and chemical structures of the genetic code (Négadi, 2007, 2008, 2009a,b).

Now, the RNA molecule is an awesome object. Recent research has shown that it has passive (information carrier or structural) as well as active (functional) roles in the innumerable biochemical processes in the living organisms. The number of kinds of RNA is growing fast. In addition to the well known three types mRNA (messenger RNA), tRNA (transfer RNA) and rRNA (ribosomal RNA), a countless number of RNAs, with an active role, have been discovered in the last decades. Some few examples of them are: snoRNAs, miRNAs, siRNAs, shRNAs), tasiRNAs, rasiRNAs, eRNAs, etc., the list is still long and certainly not closed. The growing importance of RNA over DNA has led scientists to even challenge the time-honored Central Dogma of Molecular Biology, symbolized by the sequence $DNA \rightarrow RNA \rightarrow \text{proteins}$. For example, Brosius (Brosius, 2003a) writes “On a evolutionary level one might rather suggest a restatement as $RNA \rightarrow \text{proteins} \rightarrow DNA$ ”. In another paper (Brosius, 2003b), he says that DNA and proteins, as *latecomers*, could have been subsequent “inventions” of RNA. It seems therefore that the ribonucleic acid (RNA) is a major player, compared to DNA.

We have found, in our investigations these last few years, that the atomic structure of the basic building-blocks of RNA, the four ribonucleotides, is stuffed with Fibonacci numbers, see below. The chemical formulas of the four ribonucleotides are as follows: UMP uridine monophosphate: $C_9H_{13}N_2O_9P$ (U: $C_4H_4N_2O_2$), CMP cytidine monophosphate: $C_9H_{14}N_3O_8P$ (C: $C_4H_5N_3O_1$), AMP adenosine monophosphate: $C_{10}H_{14}N_5O_7P$ (A: $C_5H_5N_5$), GMP guanosine monophosphate: $C_{10}H_{14}N_5O_8P$ (G: $C_5H_5N_5O_1$). Obviously, the symbols H, C, N, O and P refer to hydrogen (a light atom), carbon, nitrogen, oxygen and phosphorus, respectively (heavy atoms). In the parentheses we have indicated the chemical (atomic) composition of the nucleotides uracil (U), cytosine (C), adenine (A) and guanine (G), respectively. From these formulas, we have 144 atoms in the four ribonucleotides (H: 55, CNOP: 89). (One of the most important results we have obtained from the Fibonacci number 144 (see Négadi, 2013) is a mathematical form of the “*Condensation Equation*” for the formation of a ribonucleotide from the three sub-units, a nitrogenous base N, a phosphate group (8 atoms) and ribose (20 atoms) with the release of two water molecules (6 atoms). We established $N + \text{ribose} + \text{phosphate} - (2H_2O) = N + 20 + 8 - 6$. Importantly for what follows, uridine monophosphate will play a special role as it has 13 hydrogen atoms and 21 CNOP atoms (total 34). Moreover, uridine monophosphate¹ has a Hydrogen Donor Count (HDO) equal to 5 and a Hydrogen Acceptor Count (HAO) equal to 8 ($5 + 8 = 13$). Also, there are 55 hydrogen atoms in the other three ribonucleotides CMP, AMP and GMP. Note that we have here 8 Fibonacci numbers (5, 8, 13, 21, 34, 55, 89 and 144), of which 5 correspond to uridine *only*. (Note also that we have 5 for the total number of the possible nucleotides, 2 for the “interchangeable” U and T (RNA/DNA) and 3 for C, A and G, which are common to both RNA and DNA.) We have therefore eight members of the Fibonacci sequence (0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, ...) in the four ribonucleotides, with 5 of them (out of 8) in uridine only. This very fact, has led us to consider the partition “uridine and the rest”, in atom-numbers: $34 + 110 (= 34 + 2 \times 55)$. Interestingly, this latter partition is seen to obey the following Fibonacci identity $F_n + 2F_{n+1} = F_{n+3}$ for $n=9$. Using the functions defined above and adding the length-function $\lambda(n)$ of the two numbers 34 and 110 (the number of their digits), an addition which will prove quite appropriate, we have

$$A_0(34) + A_0(110) + \lambda(34) + \lambda(110) = 29 + 30 + 2 + 3 = \{59 + 2\} + 3 = 61 + 3 = 64 \quad (4)$$

¹ <http://www.hmdb.ca/metabolites/HMDB0000288>
<https://www.drugbank.ca/drugs/DB03685>

This equation gives the total number of codons, 61 (59 codons for which there is a synonymous alternative and coding for 18 amino acids, *see below about this latter number*, and 2 codons corresponding to the two singlets Tryptophane and Methionine), and the number of stops, 3. Interestingly, the above equation could be partitioned into the two following parts:

$$\{A_0(34)+\text{SPI}(110)+\Omega(110)\}=41 \quad (5)$$

$$\{a_0(110)+\lambda(110)+\lambda(110)\}=18+2+3=23 \quad (6)$$

The sub-partitions in Eqs.(5) and (6) prove to be very useful to establish the connection with our “*One Number Model*” of the genetic code, that is Eqs.(1) and (2). Let us first look at Eq.(6). This is exactly the number of digits in Eq.(1), 18 digits (1 to 9) with many multiplicities, and 5 zeros (2+3), see Eq.(3). Moreover, and this is a very nice fact which goes “like a glove”, the function $a_0(110)$ in Eq.(6) writes $a_0(110)=2+5+11$ and we see that in Eq.(1) there are 11 *even* digits, 5 *odd prime* digits and 2 *odd non-prime* digits (*total 18*). Alternatively, and interestingly, there are 11 *non-prime* digits and 7=2+5 *prime* digits: {2, 2} and {3, 5, 5, 7, 7}. The sub-partition (6) describes therefore, perfectly, Eq.(1) and the biological data reality! Now, in Eq.(5), instead of summing the various functions $a_0(34)=19$, $\text{SPI}(34)=1+7$, $\Omega(34)=2$, $\text{SPI}(110)=9$ and $\Omega(110)=3=1+1+1$, we can also consider them *separately*. Let us mention here that, except for the above mentioned (concrete) presence of Fibonacci numbers, we have also, with respect to uridine, a superimposed collection of Lucas numbers, albeit in a hidden way. Recall for example that for uridine, $A_0(34)=29$ which is the seventh Lucas number. Using the above functions and also some few mathematical identities for Fibonacci (and Lucas) numbers, we are finally led to a collection of nine numbers that we collect and arrange in decreasing order in the following sequence

$$[19, 9, 4, 3, 2, 1, 1, 1, 1] \quad (7)$$

where, in the last step, we have written 7, the fourth Lucas number, which is part of the function $\text{SPI}(34)$ above, as 3+4. Finally, applying a simple Gödel-encoding procedure (see Négadi, 2008, 2009b), we get exactly Eq.(2). This completes the demonstration of our claim: the “*galore*” of Fibonacci numbers in the RNA units, the four ribonucleotides, imply our One-Number Model of the genetic code. So, after a decade or so of (possible) interrogations, one could now understand, and maybe appreciate, that

the number, $23!$, so rich by its amazing consequences with respect to the genetic code mathematical and chemical structures, does not come from nowhere.

REFERENCES

- Brosius, J. (2003a) The contribution of RNAs and retroposition to evolutionary novelties, *Genetica*, 118, 99-116.
- Brosius, J. (2003b) Gene duplication and other evolutionary strategies: from the RNA world to the future, *Journal of Structural and Functional Genomics*, 3, 1-17.
- Négadi, T. (2007) The genetic code multiplet structure, in one number, *Symmetry: Culture and Science*, 18(2-3), 149-160.
- Négadi, T. (2008) The genetic code via Gödel encoding, *The Open Physical Chemistry Journal*, 2, 1-5.
- Négadi T. (2009a) The genetic code degeneracy and the amino acids chemical composition are connected, *NeuroQuantology*, Vol. 7, 1, 181-187 (<http://arxiv.org: q-bio.OT/0903.4131>).
- Négadi, T. (2009b) A taylor-made arithmetic model of the genetic code and applications, *Symmetry: Culture and Science*, 20(1- 4), 51-76.
- Négadi, T. (2013) The genetic code invariance: when Euler and Fibonacci meet, *Symmetry: Culture and Science*, 25(3), 261-278. ([http://arxiv.org: 1406.6092v1 \[q-bio.OT\].](http://arxiv.org: 1406.6092v1 [q-bio.OT].)).