

Received: 2003.09.22
Accepted: 2004.01.05
Published: 2004.04.01

Can the genetic code be mathematically described?

Diego L. Gonzalez

St. George School Foundation and National Research Council of Italy, Isola di San Giorgio Maggiore, Venice, Italy

Source of support: Self financed.

Summary

From a mathematical point of view, the genetic code is a surjective mapping between the set of the 64 possible three-base codons and the set of 21 elements composed of the 20 amino acids plus the Stop signal. Redundancy and degeneracy therefore follow. In analogy with the genetic code, non-power integer-number representations are also surjective mappings between sets of different cardinality and, as such, also redundant. However, none of the non-power arithmetics studied so far nor other alternative redundant representations are able to match the actual degeneracy of the genetic code. In this paper we develop a slightly more general framework that leads to the following surprising results: i) the degeneracy of the genetic code is mathematically described, ii) a new symmetry is uncovered within this degeneracy, iii) by assigning a binary string to each of the codons, their classification into definite parity classes according to the corresponding sequence of bases is made possible. This last result is particularly appealing in connection with the fact that parity coding is the basis of the simplest strategies devised for error correction in man-made digital data transmission systems.

key words: genetic code • amino acid degeneracy • parity coding • error correction

Full-text PDF: http://www.MedSciMonit.com/pub/vol_10/no_4/4184.pdf

Word count: 2375
Tables: –
Figures: 3
References: 25

Author's address: Diego L. Gonzales, St. George School Foundation and National Research Council of Italy, Isola di San Giorgio Maggiore, I-30124, Venice, Italy. e-mail: diego.gonzalez@cini.ve.cnr.it

Amino acid coding obeys a digital scheme: four distinct symbols grouped in clusters of three elements, or triplets, code for the twenty different amino acids and the Stop signal which marks the end of protein transcription [1]. This digital nature of amino acid encoding allows for the development of computational models of DNA which in turn permit the application of information theory techniques for the analysis and interpretation of long genome sequences [2]. Although quaternary number systems can be utilized to represent triplets numerically [3], efficient computational encoding is usually of a binary type, that is, it uses only two different symbols (0,1) instead of the four basic symbols T, C, A, and G associated with the four bases of the double strand of DNA (see Box 1). Mainly for this reason (but not exclusively), several methods have been proposed to convert sequences of bases into bit sequences. Almost all of them use a fixed binary representation of two bits for any base within the triplet (one bit if only the purine-pyrimidine (Y-R) character of a base in a specified position with respect to the reading frame is coded) [4–7]. However, the usual binary description represents numbers univocally, and because of this inherently injective (one-to-one) property, does not contribute very much to uncovering any mathematical order related to the weaker surjective (onto) character of the genetic code. In this sense, more interesting are binary non-power representations (see Box 1), because a given whole number can be encoded simultaneously by several binary strings: there is an essential degeneracy in the representation. A general question can thus be posed as to whether there exists a binary non-power number representation describing the degeneracy of the genetic code, or, in mathematical terms, whether there exists a binary non-power representation which defines a structural isomorphism between the two surjective applications.

To answer this question, we first note that the palindromic symmetry of non-power representations implies that in a given representation, the cardinal number of at most one subset of represented numbers with a given degeneracy can be odd (see Box 1), or, in other words, the number of whole numbers which share the same degeneracy in the representation is always even (with only one possible exception). As a consequence, the actual degeneracy of the genetic code cannot be described by any non-power integer representation because three subsets of amino acids sharing the same degeneracy are odd, i.e. there are 3 amino acids with degeneracy 6, 5 amino acids with degeneracy 4, and 9 amino acids with degeneracy 2. However, different works suggest that what matters is degeneracy inside quartets of the genetic code [7–9], that is, inside groups of 4 triplets determined by the same first two letters. This is equivalent to saying that degeneracy-6 amino acids are indeed coded by two subsets, one with degeneracy 2 and the other with degeneracy 4. In such a case it can be demonstrated that a non-power representation of the genetic code degeneracy is possible if we assign to the 6 positional base numbers of the representation the values $(x_1, x_2, \dots, x_6) = (1, 1, 2, 4, 7, 8)$. We call this particular non-power representation a DNA-like binary non-power representation (see Figure 1). The DNA-like

representation describes exactly all the actual degeneracy of the genetic code inside quartets, i.e. there are the same number of amino acids sharing a given degeneracy in the genetic code as that of whole numbers sharing the same degeneracy in the DNA-like representation, and this is valid for all degeneracy values (from 1 to 4). In fact, there are two objects (which can be whole numbers or amino acids) with degeneracy 1, 12 with degeneracy 2, 2 with degeneracy 3, and 8 with degeneracy 4 (the degeneracy-6 amino acids contribute with 3 elements to each of the degeneracy-2 and -4 families).

This fact represents *per se* a surprising mathematical result, because the probability that such an exact correspondence of degeneracy can be due to a fortunate coincidence (indeed calculated for a less restrictive case known as Rumer's transformation) is very low [10]. However, the former description will remain only a result of theoretical interest if a clear connection between the binary strings and the real triplets of bases cannot be done. For this reason, in order to find the biological meaning of the present mathematical description of the code degeneracy, we make an attempt to establish the connection between binary strings and base triplets based on the symmetry properties of both the DNA-like non-power representation and the genetic code. The results are also presented in Figure 1, where to every whole number is assigned a coded amino acid. Some of the main aspects related to this surprising correspondence between triplets and binary strings are discussed in the following:

- i) 3rd letter coding: All known versions of the genetic code possess a perfect symmetry under the exchange of C↔T in the last letter of the triplet. In Figure 1 we can observe that the same is true for strings ending with 01 or 10: all the encoded numbers represented by these kinds of strings remain invariant under the associated transformation 01↔10. Thus we may conclude that $a_6, a_5, a_4, a_3, 0, 1 - a_6, a_5, a_4, a_3, 1, 0$ exchange is equivalent to C-T exchange in the last letter of the triplet (with the encoded amino acid remaining invariant under such a transformation). By exclusion, strings ending in 00 or 11 must encode a final G or A, but observing that both degeneracy-one amino acids, W Trp and M Met, are necessarily encoded by the degeneracy-one strings, (0, 0, 0, 0, 0, 0) and (1, 1, 1, 1, 1, 1), and that both end in G, we see that the termination of strings in 00 or 11 does not suffice to distinguish triplets ending in A or G. We need to consider the parity of the strings. For a binary string, parity is defined according to the total number of 1's: an even number of 1's defines an even-parity string while an odd number of 1's defines an odd-parity string (see also Box 1). Returning to the A or G coding of the last letter of a triplet, we can assert that strings ending in 00 or 11, but with total even parity, encode a final G, while the same kind of strings with total odd parity encode a final A.
- ii) palindromic symmetry: By inverting the former rules we can conclude that the two degeneracy-3 strings represent triplets ending in T, C, and A. This descri-

Usual number representations use the power of some basis, k , weighted with digits that span the range $(0; k-1)$ [22]. This ensures that the representation is univocal (one to one). In the digital representation is used the base 10 ($k=10$) and the positional coefficients, which can take the values from 0 to 9, are placed following the rising ordered powers of 10. For example the base 10 representation of the number 15 coincides with its usual written version with the meaning,

n	$10^1=10$	$10^0=1$
15	1	5

$$15 = 1.10 + 5.1$$

The binary representation corresponds to $k=2$, so the positional values correspond to the powers of 2 and the relative coefficients can take only the values, 0 and 1 [22]. The number 15 is represented in the binary base as,

n	$2^3=8$	$2^2=4$	$2^1=2$	$2^0=1$
15	1	1	1	1

$$15 = 1.8 + 1.4 + 1.2 + 1.1$$

A generalization of the k -power number system consists in the assignment of arbitrary numbers to the positional values instead of the powers of k . For this reason this systems are sometimes called non-power representations [23]. If the positional values grow more slowly than the powers of two the representation is in general complete (all numbers from 0 to the sum of all bases are represented) but redundant (a given number can be represented by more than one binary string). An interesting example is the Fibonacci representation [23–24] for which the positional base values are successive Fibonacci numbers. Fibonacci numbers take their name from Leonardo Pisano, known also as Fibonacci, the italian mathematician who discover them. Fibonacci numbers form a series in which the n th number of the series is obtained as the sum of its two predecessors, i.e., $F_n = F_{n-2} + F_{n-1}$, with the initial condition $F_1 = 1$ and $F_2 = 1$. The Fibonacci representation of order 6 (6 bits 'word size') uses the first 6 Fibonacci numbers, 1, 1, 2, 3, 5, and 8. The number 15, for example, is represented by the following binary strings,

n	8	5	3	2	1	1	8	5	3	2	1	1	8	5	3	2	1	1
15	1	1	1	1	0	0	1	1	0	0	1	1	1	0	1	1	1	1

$$15 = 1.8+1.5+1.2 = 1.8+1.5+1.1+1.1 = 1.8+1.3+1.2+1.1+1.1$$

We have highlighted in colour the concept of parity of a binary string: even strings are in red, odd strings are green. Parity is defined taking into consideration the number of 1's in a given string: an even number of 1's gives an even string and an odd number of 1's an odd string. The order-6 Fibonacci representation shares some properties with the genetic code, for example, is redundant (the number 15 is represented by 3 binary strings) and there are exactly 21 numbers represented by 64 binary strings. However, the degeneracy of the order-6 Fibonacci coding do not describes the actual degeneracy observed in the genetic code (nor other alternative redundant numeration systems [25]).

In a more general context, non-power number representations possess a general property: degeneracy is a palindromic function of the represented number; the numbers r , and $R-r$, where R is the maximum integer that can be represented, share the same degeneracy. Palindromically related pairs are numerically represented by a complement to 1 operation: the palindrome of a given string is obtained replacing simultaneously 0 by 1 and 1 by 0. A consequence of the palindromic property is that the subset of represented numbers with a given degeneracy has a cardinal number even except for the case in which $(R+1)$ is odd. In such a case, the degeneracy of the central number, $R/2$, is shared by an odd number of represented numbers.

Box 1. Integer number representations.

bes well the I Ile assignment, but not the Stop signal. The Stop signal is the only one with degeneracy less than or equal to 4 that has triplets with a different second letter. In fact, the triplet TGA presents some odd characteristics. For example, it codes for different amino acids in different versions of the code and, in some cases, for a 21st amino acid, selenocysteine [11–12]. Thus we can think that this assignation in the standard code is a kind of symmetry breaking of a more symmetric template, as has also been suggested in other contexts [13–14]. We

can observe that in a particular version of the genetic code, the *euplotid nuclear*, the triplet TGA represents C Cys, which becomes one of the degeneracy-3 amino acids (instead of the Stop signal) being represented by three triplets ending in T, C and A, as implied by the DNA-like representation. For this reason, the *euplotid nuclear* version of the code is considered, in the present approach, the general symmetric template on which different symmetry-breaking alterations describe the other versions, and in particular the standard one (which differs from

This copy is for personal use only - distribution prohibited.

#	8 7 4 2 1 1						8 7 4 2 1 1						8 7 4 2 1 1						8 7 4 2 1 1						#						
	8	7	4	2	1	1	8	7	4	2	1	1	8	7	4	2	1	1	8	7	4	2	1	1							
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	23
1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	1	1	1	0	1	22	
2	0	0	0	1	0	0	0	0	0	1	1	0	0	0	0	1	1	0	0	0	0	1	1	0	1	1	0	1	1	21	
3	0	0	0	1	1	0	0	0	0	1	0	1	0	0	0	1	0	1	0	0	0	1	0	1	1	1	0	0	1	20	
4	0	0	1	0	0	0	0	0	0	1	1	1	0	0	0	1	1	1	0	0	0	1	0	0	1	1	0	1	1	19	
5	0	0	1	0	1	0	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	1	1	0	1	0	18	
6	0	0	1	1	0	0	0	0	1	0	1	1	0	0	1	0	1	1	0	0	1	0	1	1	1	1	0	1	1	17	
7	0	0	1	1	1	0	0	0	1	1	0	1	0	1	0	0	0	0	0	1	0	0	1	0	1	1	0	0	1	16	
8	1	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0	1	0	1	0	0	0	1	1	0	1	1	1	15	
9	1	0	0	0	1	0	1	0	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	1	0	1	14	
10	1	0	0	1	0	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	0	1	0	0	1	0	1	1	13	
11	1	0	0	1	1	0	1	0	0	1	0	1	0	1	1	0	0	0	1	0	1	0	0	1	0	1	1	0	0	12	

Figure 1. DNA-like binary representation. Representation of the first 24 whole numbers (shown in the outside columns) in the DNA-like binary non-power representation. The degeneracy of the genetic code inside quartets is exactly described by the numerical representation; for a given whole number, degeneracy corresponds to the number of length-6 binary strings horizontally filled. In the central columns are represented the degeneracy number and the pair of palindromic amino acids that can be assigned to the binary strings on the basis of symmetry considerations for the *euploid nuclear* version of the genetic code. A particular degeneracy-conserving symmetry, the *palindromic symmetry*, is evident as a horizontal mirror symmetry between the two groups of strings (an imaginary mirror can be placed vertically between the left and right groups). At the arithmetical level, the palindrome of a string is obtained by means of the exchange of 0 by 1 and vice versa. Because of this perfect symmetry the absolute position of a given amino acid in the table is not completely determined: single amino acids can be exchanged inside a palindromic pair without affecting the degeneracy number. String parity is also represented: even-parity strings are shown in red and odd-parity strings in green (i.e. an even or odd number of 1's in the corresponding string). Observe that palindromic symmetry also preserves parity (see Box 1).

This copy is for personal use only - distribution prohibited.

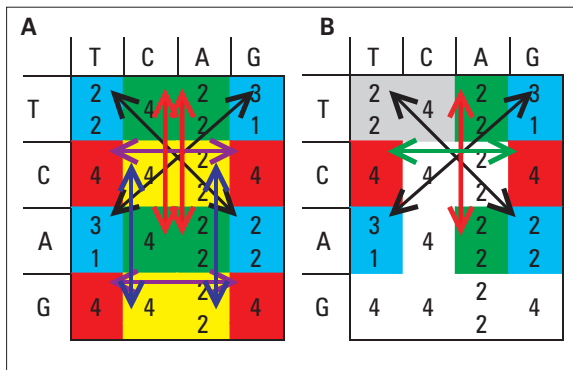


Figure 2. (A). Palindromic symmetry. All quartets (groups of four codons sharing their two first letters) are related in pairs by the palindromic transformation. Arrows of the same color indicate a common operation at the triplet level. The complete set of transformations can be written in compact form as:

	1 st letter	2 nd letter	3 rd letter
Transformation	[T↔A; C; G]	[T↔G]	[T↔C; A; G]
Transformation	[T↔A; C↔G]	[A; C]	[T↔C; A; G]

which has not been previously noted (some similar mapping is represented by the Rumer's transformation which do not preserve degeneracy) [7]. Observe that the letter T is the only one that is transformed in all triplet positions. The symmetry corresponds exactly to the *Euplotid nuclear* version of the genetic code. The standard code can be viewed as a symmetry breaking at the level of the codon TGA, which codes in that case for the Stop signal instead of Cysteine. (B) Site variations and palindromic transformation. The quartets involved in some variation of the encoded amino acids or in the Stop triplets among the different versions of the genetic code are highlighted in color [15]. There are 8 variable quartets related by the palindromic transformation. The only exceptions are shown in gray: TCN, for which there is reported a variation in the triplet TCA [16], and TTN, for which there are no reported variations. Similarly, the invariant quadruplets (with the exception of TCN) are also related by the palindromic transformation.

the *euplotid nuclear* one only in the TGA triplet assignment). We can observe that in the *euplotid nuclear* version, this symmetry is represented by transformations between quartets (groups of four triplets) of the genetic code which preserve the coding degeneracy. This fact is in some sense complementary of the before mentioned Rumer's transformation [7] describing a mapping between quartets which break the coding degeneracy. The noticeable fact in the present approach is that this coding degeneracy-preserving symmetry is described at the mathematical level by a simple and concise rule: two symmetric triplets are represented by binary strings related by the complement to 1 operation (replace any 0/1 in one string by 1/0 in the symmetric one). We call this remarkable symmetry *palindromic symmetry*. The term palindromic symmetry refers here to the mathematical properties of the binary strings representing triplets and their expression at the level of the genetic code structure, not to the actual sequence of bases along the DNA

	T	C	A	G	
T	TTT Phe	TCT Ser	TAT Tyr	TGT Cys	T
	TTC Phe	TCC Ser	TAC Tyr	TGC Cys	C
	TTA Leu	TCA Ser	TAA Stop	TGA Cys	A
C	TTG Leu	TGG Ser	TAG Stop	TGG Trp	G
	CTT Leu	CCT Pro	CAT His	CGT Arg	T
	CTC Leu	CCC Pro	CAT His	CGC Arg	C
A	CTA Leu	CCA Pro	CAA Gln	CGA Arg	A
	CTG Leu	CCG Pro	CAG Gln	CGG Arg	G
	ATT Ile	ACT Thr	AAT Asn	AGT Ser	T
G	ATC Ile	ACC Thr	AAC Asn	AGC Ser	C
	ATA Ile	ACA Thr	AAA Lys	AGA Arg	A
	ATG Met	ACG Thr	AAG Lys	AGG Arg	G
G	GTT Val	GCT Ala	GAT Asp	GGT Gly	T
	GTC Val	GCC Ala	GAC Asp	GGC Gly	C
	GTA Val	GCA Ala	GAA Glu	GGA Gly	A
	GTG Val	GCG Ala	GAG Glu	GGG Gly	G

Figure 3. Parity marking of triplets. Using the same color code as in Figure 1, even- and odd-parity triplets are shown. Triplets ending in A or G are, respectively, odd and even. For triplets ending in C or T, the second letter of the triplet determines parity: T or G determines odd parity, while C or A determines even parity. Observe that there is a perfect correspondence with the parity of the binary strings reported in Figure 1.

chain. In fact, if a string representing a particular whole number p is palindromic to another string, the latter represents the whole number $23-p$. It can be observed in Figure 1 that the degeneracy corresponding to a given whole number p is the same as that corresponding to the complementary number $23-p$; thus, the degeneracy list is unchanged if we read it from 0 to 23 or from 23 to 0, that is, it is a palindromic function of the whole number represented. Regarding the genetic code, the total set of palindromic transformations for the *euplotid nuclear* version of the code is represented in graphical form for the two first letters of triplets in Figure 2A. As the palindromic transformation does not change parity at the string level, the result of its application at the level of the third letter of triplets is trivial: C and T are exchanged, while A and G remain invariant. Many interesting properties of the representation will be described in a more extensive work (*DNA-like number representations and the internal mathematical structure of the genetic code*, D. L. Gonzalez, in preparation). Here we remark only two other facts related to palindromic symmetry:

- i) the position of variable triplets in the different versions of the code: in Figure 2B are represented the triplet variations for all known versions of the genetic code taking as a general template the *euplotid nuclear* one (observe that the position of variable sites represents an absolute reference frame) [15]. Remarkably, variations in assignment occur in quartets related by the palindromic transformation which define exactly two halves of the genetic code composed of 8 quartets: one half prone to codon reassignments and the other invariant. Only 1 out of the 24 variations reported breaks this striking symmetry [16];
- ii) parity rules: the set of palindromic transformations is compatible with the following parity rule for triplets

This copy is for personal use only - distribution prohibited.

ending in T or C: triplets ending in T or C and with total even parity have a C or an A as the second letter; those with total odd parity have a T or G as the second letter. This rule complements the former one for triplets ending on A or G and shows that all triplets of the code are parity marked: if the triplets end in A or G, the mark is coincident with that letter, the string being even for a final G and odd for a final A; for triplets ending in C or T, the parity is instead coded on the second letter of the triplet (see Figure 3).

Again, as in the case of palindromic symmetry, we found that a very simple and concise mathematical property of the binary strings, i.e. parity, is related to a definite property of the triplets in terms of their base arrangement. It seems that triplets follow a very structured organization which is based on these simple mathematical rules established at the level of their equivalent binary non-power representation.

We have determined a coding system based on a particular non-power number representation which describes exactly and in a natural way the actual genetic code degeneracy. Moreover, this system allows for a binary representation of triplet structure. The binary code is not fixed, but it is context dependent: binary coding of a particular base depends on the other letters of the triplet and on its position relative to the reading frame. This fact, together with the parity 'marking' of triplets by means of their second or last letter, strongly suggests the presence of an error-correction mechanism based on parity checking [17]. This possibility has indeed been suggested [18], and at least from a linear point of view, investigated [19]. Remarkably, a parity-coding strategy has also been proposed recently for the selection of complementary bases in the double strand of DNA [20]. Our approach strongly suggests that there exists a correction-error mechanism acting also 'along' the strands and that such a mechanism could contribute to the accuracy of protein synthesis. In fact, different from the parity coding of complementary pairs, which has meaning mainly in an evolutionary context, our hypothesized mechanism could be active in the complex transcription-translation synthesis path. In this sense it could be a concurrent factor for the very low amino acid error rates in protein synthesis and also for the different rates of error depending on the positions of nucleotides in codons relative to the reading frame (error rate asymmetry). This fact may be of particular importance in medicine, because a defect in the hypothesized error correction system could be responsible for a higher rate of error or the impossibility to correct specific errors associated to specific diseases. Moreover, a simple protein of 100 amino acids in length can be coded in 3^{100} , i.e. approximately 10^{47} , different ways. How and why is picked-out a particular sequence of nucleotides in this sea of possibilities for the coding of a specific protein? The present approach offers an opportunity to explore in depth some of these organizational aspects of biological information coding in DNA. The most natural suggestion in this sense is to analyze the statistical properties of real coding sequences of DNA assigning to triplets the non-trivial coding of Figure 1. Also other

simplified statistics can be performed, for example analyzing only the parity properties of triplets (only one coding bit for triplet).

From the theoretical point of view, different theories have been developed to explain observed or inferred code regularities [20]. In some cases these can be associated with physico-chemical properties of the molecules involved in the amino acid coding and decoding processes [8,13,21]. It remains striking, however, that different fundamental properties of the genetic code, such as degeneracy distribution, and also unexpected hidden properties, such as the palindromic symmetry and the parity marking of triplets presented here, reflect a strong mathematical order which is accurately described by means of one of the most elementary operations at the root of mathematics: number representation.

REFERENCES:

1. Brown, TA: Genome. Second Edition, BIOS Scientific Publishers, Oxford, 2002
2. Yockey HP: Information Theory and Molecular Biology. Cambridge University Press, Cambridge, New York, 1992
3. Jimenez-Montañó MA, de la Mora-Basañez CR, Poschel T: The hypercube structure of the genetic code explains conservative and non-conservative aminoacid substitutions *in vivo* and *in vitro*. Biosystems, 1996; 39: 117-25
4. Jimenez-Montañó MA, de la Mora-Basañez, CR: The genetic code as a six-dimensional boolean hypercube. In: Abstracts of Proc Soc Math Biol, Annual Meeting, July 23–26, U.C, Berkeley, CA, 1992
5. Karasev VA, Sorokin SG: Topological structure of the genetic code. Russ J Genet, 1997; 33: 622-28
6. Klump HH: The physical bases of the genetic code: the choice between speed and precision. Archv Biochem Biophys, 1993; 301: 207-9
7. Rumer YB: About the codon's systematization in the genetic code (in Russian), Proc Acad Sci U.S.S.R. (Doklady), 1966; 167: 1393
8. Lehmann J: Physico-chemical constraints connected with the coding properties of the genetic system. J Theor Biol, 2002; 202: 129-44
9. Shcherbak VI: The symmetrical architecture of the genetic code systematization principle. J Theor Biol, 1993; 162: 395-98
10. A calculation regarding the possibility of casual generation of the so-called Rumer's transformation, which presents some analogies with the present approach in the sense that the palindromic transformation (see text) also connects two groups of 8 quadruplets, gives a probability of less than 3.09^{-32} ; Zhaxybayeva O, Statistical estimation of Rumer's transformation of the universal genetic code. ISSOL'96, Orleans, France, 1996
11. Zinoni F, Birkmann A, Stadtman TC, Bock A: Nucleotide sequence and expression of the selenocysteine-containing polypeptide of formate dehydrogenase (formate-hydrogen-lyase-linked) from *Escherichia coli*. Proc Nat Acad Sci, 1986; 83: 4650
12. Atkins JF, Gesteland RF: Selenocysteine, the 21st amino acid. Nature, 2000; 407: 463-65
13. Jimenez-Montañó MA: Protein evolution drives the evolution of the genetic code and vice versa. Biosystems, 1999; 54: 47-64
14. Hornos JEM, Hornos YMM: Algebraic model for the evolution of the genetic code. Phys Rev Lett, 1993; 71: 4401-4
15. Knight RD, Freeland SJ, Landweber LF: Rewiring the Keyboard: Evolvability of the Genetic Code. Nature Reviews – Genetics, 2001; 2: 49-58
16. Kuck U, Jekosch K, Holtzamer P: DNA sequence analysis of the complete mitochondrial genome of the green alga *Scenedesmus obliquus*: evidence for UAG being a leucine and UCA being a non-sense codon. Gene, 2000; 253: 13-18
17. Sweeney P: Error Control Coding. Wiley, New York, 2002
18. Forsdyke D: Are introns in-series error-detecting sequences, J Theor Biol, 1981; 93: 861-66

19. Liebovitch LS, Tao Yi, Todorov AT, Levine L: Is there an error-correcting code in the base sequence in DNA?, *Biophys J*, 1996; 71: 1539-44
20. Donall A, Mac Donail: A parity code interpretation of nucleotide alphabet composition. *Chem Commun*, 2002; 18: 2062-63
21. Di Giulio M: On the origin of the genetic code. *J Theor Biol*, 1997; 187: 573-81
22. O Ore: *Number Theory and its History*. Dover Publications, New York, 1988
23. Wolfram SA: *New Kind of Science*, Wolfram Media, Illinois, 2002
24. Zeckendorf E: Représentation des nombres naturels par un somme de nombres de Fibonacci ou de nombres de Lucas. *Bulletin de la Société Royale des Sciences de Liege*, 1972; 41: 179-82
25. Knuth DE: *The Art of Computer Programming, volume 2, Seminumerical Algorithms*, third edition. Addison Wesley, Reading, Massachusetts, 1997

This copy is for personal use only - distribution prohibited.

Index Copernicus

Global Scientific Information Systems for Scientists by Scientists

www.IndexCopernicus.com



TM
INDEX
COPERNICUS
INTERNATIONAL

EVALUATION & BENCHMARKING

PROFILED INFORMATION

NETWORKING & COOPERATION

VIRTUAL RESEARCH GROUPS

GRANTS

PATENTS

CLINICAL TRIALS

JOBS

STRATEGIC & FINANCIAL DECISIONS

Index Copernicus integrates

IC Journal Master List

Scientific literature database, including abstracts, full text, and journal ranking. Instructions for authors available from selected journals.

IC Conferences

Effective search tool for worldwide medical conferences and local meetings.

IC Scientists

Effective search tool for collaborators worldwide. Provides easy global networking for scientists. C.V.'s and dossiers on selected scientists available. Increase your professional visibility.

IC Patents

Provides information on patent registration process, patent offices and other legal issues. Provides links to companies that may want to license or purchase a patent.

IC Grant Awareness

Need grant assistance? Step-by-step information on how to apply for a grant. Provides a list of grant institutions and their requirements.

IC Virtual Research Groups [VRG]

Web-based complete research environment which enables researchers to work on one project from distant locations. VRG provides:

- ④ customizable and individually self-tailored electronic research protocols and data capture tools,
- ④ statistical analysis and report creation tools,
- ④ profiled information on literature, publications, grants and patents related to the research project,
- ④ administration tools.

IC Lab & Clinical Trial Register

Provides list of on-going laboratory or clinical trials, including research summaries and calls for co-investigators.