PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON MATHEMATICS AND ENGINEERING TECHNIQUES IN MEDICINE AND BIOLOGICAL SCIENCES

METMBS'03

Editors:
Faramarz Valafar & Homayoun Valafar

Associate Editors:

Roman I. Andrushkiw, H. R. Arabnia, Syed Haque, Matthew He, Mariofanna Milanova and Shouzhong Xiao

Las Vegas, Nevada, USA
June 23-26, 2003

The Biperiodic Table and Attributive Conception of Genetic Code. A Problem of Unification Bases of Biological Languages

S. Petoukhov
Department of Biomechanics, Mechanical Engineering Research Institute, Russian Academy of Sciences, 4, Malyi Haritonievskii per., Moscow, Russia

Abstract

The author puts forward an attributive conception of genetic language. It discovers existence of three binary sub-alphabets of genetic code according to three kinds of biochemical attributes. The "elementary" four-letter alphabet of genetic code comprises these sub-alphabets. From this viewpoint, each genetic message includes three parallel texts of binary types as in computer coding. The author constructs octet biperiodic table of genetic code with interesting mathematical peculiarities. This table of 64 genetic triplets is similar to a famous table of 64 hexagrams from the Ancient Chinese "The Book of Changes" unexpectedly. Each of 64 triplets gets its natural individual number from 0 till 63 according to this attributive conception. The parallel presence of ensembles of the binary sub-alphabets at different fields of information physiology is marked. The author supposes that ensembles of binary sub-alphabets play a role of unification (unitized) base for all or for majority of biological languages.

Keywords: attribute, genetic code, language, unification.

From the information viewpoint, living organisms are information essences. They live due to receiving of the genetic information from the ancestors, and they exist to transfer the genetic information to their descendants. In the biological literature it is possible quite often to meet the statement that living organisms are the texts since a molecular level of their organization. Just from the information-hereditary viewpoint all living organisms are unified wonderfully: all of them have identical bases of system of genetic coding.

For realization of the genetic messages, which encode sequences of amino acids in proteins, all kinds of organisms utilize in their molecules of heredity DNA (and RNA – ribonucleic acid) the "alphabet" consisting of only four "letters" or nitrogenous bases: adenine (A), cytosine (C), guanine (G), thymine (T) {or uracil (U) in RNA}. Linear sequences of these four letters on strings of molecules of heredity (DNA and RNA) contain the genetic information for protein synthesis in all living bodies - from bacteria up to a whale and from worm up to a bird and man. The given set of four letters is usually considered as the elementary alphabet of a genetic code. The modern science does not know why the alphabet of genetic language has four letters (it could have any other number of the letters in principle) and why just these four nitrogenous bases are chosen by nature as elements of the genetic alphabet from billions possible chemical compounds.

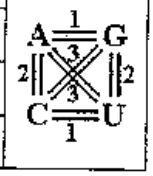
The author paid attention to the fact that these four nitrogenous bases represent specific poly-nuclear constructions with the special biochemical properties. The set of these four constructions has on itself a substantial system of attributes for uniting and

distinguishing of its letters (or, more precisely, the set has pairs of "attribute – anti-attribute"). This system of attributes divides the four-letter alphabet into various pairs of letters, which are equivalent from a viewpoint of one of these attributes or its absence: 1) C=U и A=G (according to attributes: "pyrimidine" or "non-pyrimidine", that is purine); 2) A=C и G=U (according to attributes: amino-mutating or non-amino-mutating under action of nitrous acid HNO_2); 3) C=G и A=U (according to attributes: three or two hydrogen bonds are materialized in these complementary pairs).

Let's number these attributes by numbers N=1,2,3 and let's ascribe to each of four letters of code the symbol "1_N" in case of presence at it of an attribute under number "N" and the symbol "0_N" in case of its absence. In result we receive representation of the four-alphabetic alphabet of a code in system of its three "binary sub-alphabets to attributes " (see Table 1). The four-letter alphabet of a code is curtailed into the two-letter alphabet on the basis of each kind of attributes. For example, to the first attribute we have (instead of the four-letter alphabet) the alphabet from two letters 0_1 and 1_1, which the author names as "the binary sub-alphabet to a first attribute".

Table 1. Three binary sub-alphabets according to attributes

ATTRIBUTE	G	A	U	C
1) Belonging to pyrimidine class (with one ring in the molecule)	0_1	0_1	1_1	1_1
2) Amino-mutating (or special location of NH_2 in molecular ring)	0_2	1_2	0_2	1_2
3) Belonging to complementary pair with three hydrogen bonds	1_3	0_3	0_3	1_3



Accordingly, the genetic text as a sequence of four letters of genetic language is submitted by three parallel and various sequences of zero and unit (such binary sequences are used at a storage and transfer of the information in computers). It illustrates a following example of an initial piece of a genetic sequence for insulin (see Table 2). Figuratively speaking, the genetic message appears as a bunch of parallel texts in three different languages, and genetic sequences have a property of poly-languages. This property was missed earlier from view.

Table 2. Binary sequences in a gene of insulin

Number of attribute	Sequence of "letters" in a gene of insulin
	A U G G G C A U C G U U G A A C A G U G U .....
№ 1	0_1 1_1 0_1 0_1 0_1 1_1 0_1 1_1 1_1 0_1 1_1 1_1 0_1 0_1 0_1 1_1 0_1 0_1 1_1 0_1 1_1 .....
№ 2	1_2 0_2 0_2 0_2 0_2 1_2 1_2 0_2 1_2 0_2 0_2 0_2 0_2 1_2 1_2 1_2 1_2 0_2 0_2 0_2 0_2 .....
№ 3	0_3 0_3 1_3 1_3 1_3 1_3 0_3 0_3 1_3 1_3 0_3 0_3 1_3 0_3 0_3 1_3 0_3 1_3 0_3 1_3 0_3 .....

Each letter of the genetic language in view of its meaning in three binary sub-alphabets can be expressed by a code vector in three-dimensional Cartesian space. It is well known that a set of three letters of a genetic language forms a triplet (or a codon), which encodes one of 20 amino acids or punctuation marks of protein synthesis. Due to three binary sub-alphabets, any genetic triplet can be presented by a set of three code vectors or by numerical matrix (3x3). Generally speaking, each from these parallel information texts, based on objective biochemical attributes, can provide its own subsection of protein synthesis. It's very probable, that a process of protein synthesis is divided to relative independent subsections connected with these sub-alphabets.

The ensemble of the binary sub-alphabets of genetic language (additionally represented by the author in a form of

dichotomic trees of attributes of Yin-Yang type) is new system, not investigated earlier in biology. And the "elementary" four-letters alphabet of a code is not so elementary. These sub-alphabets allow renumbering by integers from 0 up to 63 all 64 triplets, which encode 20 amino-acids and punctuations of protein synthesis. On this basis the author has constructed a substantial octet table of genetic code with fractal properties and with double periodicity – in vertical and in horizontal directions (see Table 3) [1, 2]. Unexpectedly this bi-periodic table of genetic code has a close connection with famous table of 64 hexagrams of ancient Chinese "The Book of Changes". The review of a history of attempts of other authors to find connection between "The Book of Changes " and genetic code was published in [2, Chapter 5].

More precisely, due to attributes, which transform the four-letter alphabet of genetic language into the binary sub-alphabets, internal structure and well-ordering of the set of 64 genetic triplets set are revealed. Octet bi-periodic table of codons (Table 3) demonstrates it. All triplets are renumbered by a natural manner and are located in tabular system with double – on a horizontal and vertical – periodicity and with secretive fractal structure (described in [2] in detail).

Table 3. The bi-periodic table of genetic triplets

№	111 CHYAN	110 TUI	101 LI	100 CHEN	011 HSUN	010 KAN	001 KEN	000 KUN
111 CHYAN	CCC 63	CCA 62	CAC 61	CAA 60	ACC 59	ACA 58	AAC 57	AAA 56
110 TUI	CCU 55	CCG 54	CAU 53	CAG 52	ACU 51	ACG 50	AAU 49	AAG 48
101 LI	CUC 47	CUA 46	CGC 45	CGA 44	AUC 43	AUA 42	AGC 41	AGA 40
011 HSUN	UCC 31	UCA 30	UAC 29	UAA 28	GCC 27	GCA 26	GAC 25	GAA 24
100 CHEN	CUU 39	CUG 38	CGU 37	CGG 36	AUU 35	AUG 34	AGU 33	AGG 32
010 KAN	UCU 23	UCG 22	UAU 21	UAG 20	GCU 19	GCG 18	GAU 17	GAG 16
001 KEN	UUC 15	UUA 14	UGC 13	UGA 12	GUC 11	GUA 10	GGC 9	GGA 8
000 KUN	UUU 7	UUG 6	UGU 5	UGG 4	GUU 3	GUG 2	GGU 1	GGG 0

All 64 triplets are submitted in 64 cells of this table. Any column consists of 8 triplets equivalent of each other from the viewpoint of the binary sub-alphabet of the first attribute from Table 1. In other words, triplets of one column are indiscernible from each other in relation of positions engaged in them by pyrimidines and purines. For example, in second column of Table 3 (from the left), two pyrimidines (C or U) are on first two positions of all its triplets and one purine (A or G) is on its third position. According to Table 1, binary symbol 1_1 corresponds to pyrimidines, and binary symbol 0_1 corresponds to purines. In this connection, each column together with all its eight triplets receives an individual binary-numerical symbol from the following series of eight three-digit numbers: $0_10_10_1$, $0_10_11_1$, $0_11_10_1$, $0_11_11_1$, $1_10_10_1$, $1_10_11_1$, $1_11_10_1$, $1_11_11_1$ (in the Table 2 the indexes are omitted). While translating from binary in decimal numeration system, this series will be transformed to a series of integers 0, 1, 2, 3, 4, 5, 6, 7. In our

Table 3, all columns with their families of triplets are put in descending order under these numbers. It is obvious, that in a binary system it is possible to use not only symbols of "0" and "1", but also, for example (as it was made in Ancient Chinese "The Book of Changes"), a symbol of broken line instead of "0" and a symbol of unbroken line instead of "1". According to Chinese manner, these lines can be put in vertical trigrams which should be read from below upwards (see Table 3). Then each from eight tabular families of triplets will receive the appropriate symbol of Ancient Chinese trigram, given in the Table 3 with its Chinese name.

These octet families of triplets are known in a history of molecular genetics due to the Wittmann's work [3] on fragmentation of all set of triplets into eight families in connection with property of amino-mutating of nitrogenous bases in RNA, i.e., in connection with a second attribute from Table 1. Table 3, offered by us, represents an association of these Wittmann's octets in a family of eight tabular columns with their ordering according to the natural binary numeration. It is essential, that we construct Table 3 on the basis of real biochemical materials.

Each row of the Table 3 has also the binary number connected to an attribute of amino-mutating (an attribute №2 in Table 1) in triplets of a line. All eight triplets of each row are equivalent each other from the viewpoint of the binary sub-alphabet connected with this second attribute of Table 1. For example, its second row (from above) is characterized by a symbol $1_2 1_2 0_2$, i.e. all triplets of the row have identical quantity and order of the letters with these binary symbols 1_2 and 0_2. (In Table 3 the binary symbols are printed without indexes, but the symbols of rows are underlined and are printed with italics to distinguish them from symbols of columns). The sequence of rows in Table 3 is built in the decreasing order of their natural binary numbers and this order is broken only for two average rows, which are rearranged by their places in connection with biochemical features of Wittmann's octets.

In Table 3 each triplet has coordinates from a symbol of its row and symbol of its column. Both of these symbols can be written out jointly as a single whole six-digit binary number (for example, let us accept, that coordinate three-digit number of a tabular row goes in the beginning of this six-digit number, and coordinate number of a column goes in its end). Then all triplets will be renumbered with six-digit binary numbers. At translation of these binary numbers into decimal numeration, well-ordered series of numbers from 0 up to 63 is formed. All 64 triplets in the appropriate cells of Table 3 are numerated by numbers of this series. In this natural system of ordering of triplets, numerical invariant exists, which characterizes all pairs of "codon-anticodon" (codon and anticodon refer to triplets, which are located opposite each other in pairs of complementary strings of molecules DNA or RNA). This invariant is a sum of coordinate numbers of any codon and its anticodon, and this sum is equal to 63 always. And, on the contrary, if the sum of coordinate numbers of any two triplets in Table 3 is not equal to 63, these triplets are not a pair of "codon-anticodon". For example, codon CUA and its anticodon GAU have coordinate numbers 46 and 17 accordingly, and their sum is equal to 63. In addition to this, each codon and its anticodon are located in cells, which are inverse-symmetrical in relation of the tabular center.

At use for coordinates of triplets in Table 3 not digital but Chinese trigram forms, each triplet receives an individual coordinate symbol as an Ancient Chinese hexagram: a pile of six broken and unbroken lines. This vertical set of six lines for any triplet has at its bottom the coordinate trigram of tabular row of the triplet and has on top a coordinate trigram of its tabular column. At replacement of each triplet by its coordinate hexagram, Table 3 is transformed into the table, which coincides

wonderfully with the historically famous table of 64 hexagrams from "The Book of Changes" (Table 4) in the order of *Fu-Xi*. This book is a unique monument of world culture and philosophical thought, the first one in the row of sacred books of the East (see, for example, [4]).

According to statements of Chinese sources during thousands years, this ancient table is general natural archetype. It is very interesting that the biochemical system of a genetic code is connected with this Ancient Chinese symbolical system.

Table 4. A square arrangement of hexagrams according to Fu-Xi's order from "The Book of Changes" with supplement for each hexagram of its binary-numerical value and its equivalent in a decimal numeration

TOP→ / BOTTOM↓	111 CHYAN	110 TUI	101 LI	100 CHEN	011 HSUN	010 KAN	001 KEN	000 KUN
111 CHYAN	111111 63	111110 62	111101 61	111100 60	111011 59	111010 58	111001 57	111000 56
110 TUI	110111 55	110110 54	110101 53	110100 52	110011 51	110010 50	110001 49	110000 48
101 LI	101111 47	101110 46	101101 45	101100 44	101011 43	101010 42	101001 41	101000 40
100 CHEN	100111 39	100110 38	100101 37	100100 36	100011 35	100010 34	100001 33	100000 32
011 HSUN	011111 31	011110 30	011101 29	011100 28	011011 27	011010 26	011001 25	011000 24
010 KAN	010111 23	010110 22	010101 21	010100 20	010011 19	010010 18	010001 17	010000 16
001 KEN	001111 15	001110 14	001101 13	001100 12	001011 11	001010 10	001001 9	001000 8
000 KUN	000111 7	000110 6	000101 5	000100 4	000011 3	000010 2	000001 1	000000 0

The bi-periodic table 3 has interesting mathematical properties. For one example only, let's recollect a third attribute of Table 1, according to which the pair of complementary nitrogenous bases C and G has three hydrogen bonds, and other pair of complementary bases A and U has two hydrogen bonds. Let us replace each triplet in Table 3 by the sum of these numbers of its hydrogen bonds, considering C=G=3, A=U=2. For example, the triplet CAU will be replaced by number 7 (=3+2+2). In result the octet numerical matrix, submitted in the Table 5, will be produced. All rows and all columns of this matrix differ from each other by sequences of their numbers. But the sums of all hydrogen bonds in cells of each row and of each column are identical each other and are equal to that number 60. (This number is the basis of the universal 60-nary cyclic system, on which many Chinese systems were constructed: the 60-years' Chinese calendar, 60 tune keys of traditional Chinese music, etc). Those numbers 6, 7, 8 and 9, which symbolized young and old Yin and Yang in Ancient China, appear in cells of the given matrix only.

Table 5. Numeric matrix of the biperiodic table from viewpoint of hydrogen bonds

	111 CHYAN	110 TUI	101 LI	100 CHEN	011 HSUN	010 KAN	001 KEN	000 KUN	Σ
111 CHYAN	9	8	8	7	8	7	7	6	60
110 TUI	8	9	7	8	7	8	6	7	60
101 LI	8	7	9	8	7	6	8	7	60
011 HSUN	8	7	7	6	9	8	8	7	60
100 CHEN	7	8	8	9	6	7	7	8	60
010 KAN	7	8	6	7	8	9	7	8	60
001 KEN	7	6	8	7	8	7	9	8	60
000 KUN	6	7	7	8	7	8	8	9	60
Σ	60	60	60	60	60	60	60	60	480

A more profound consideration of the known data about genetic system from the viewpoint of the binary sub-alphabets and the bi-periodic table of genetic code (in particular, on specificity of tabular locations of 20 amino acids and of punctuation marks of protein synthesis which are encoded by 64 triplets) reveals the whole hierarchy of binary systems of Yin-Yang type at different levels of genetic system. The bi-periodic table of genetic code is connected with certain dichotomy Yin-Yang trees of binary sub-alphabet elements. It gives new opportunities for the analysis of genetic phenomena on the basis of the theory of fractals and synergetic. Symmetrological analysis of described genetic structures in connection with ideas of unitary symmetries is very interesting also [2].

Together with "binary sub-alphabets to attributes", the biperiodic table is used for the profound understanding of the genetic texts; for analysis of structural features of hierarchical system of genetic coding; for study of interrelations in system of 64 triplets and in system of 20 amino-acids coded by them; for research of questions of biological evolution, etc. The author examines a task of modeling of biological objects and information packages by means of the theory of structures over the ensembles of the binary sub-alphabets; a connection of binary sub-alphabetic genetic languages with specificity of living substance and mathematical cyclic codes; formal structural parallelism between the system of pairs "attribute - anti-attribute" in genetic alphabet and a set of quarks in physics of elementary particles. The author studies also possible variants of using of genetic codes in organisms for coding not amino-acids but for coding of other essences, for example, of biological cycles.

In view of these materials, a *"binary-attributive theory of genetic code"* (or, more briefly, "attributive theory") is developing by the author. This theory supposes that binary-opposite attributes play most important role in information aspects of genetics – a role of main information units. In our opinion, nitrogenous bases A, C, G, U (T), which are considered as main information units by most authors ordinarily, play an auxiliary (accessory) role of vehicles (bearer) for these main personages.

This viewpoint contradicts a traditional belief of many authors that these nitrogenous bases (or nucleotides) are main information units in genetic messages. According to the attributive theory, the main attention in information genetic problems should be concentrated to analyze of sequences of binary-opposite attributes in genetic messages without habitual exclusive attention to sequences of nucleotides only. A set of genetic molecules, which was the most interesting for biochemists ordinarily and which determined a name "molecular genetics", is absolutely another than a set of attributes, importance of which is declared by the attributive theory. Perhaps, "attributive genetics" will be originated in future to study a connection between these attributes of genetic molecules and attributes of higher levels of biological organisms.

This initial theory gives an opportunity for many new investigations. An interesting example of it is investigations by He [5], who discovered a connection between attributive description of genetic system and Fibonacci numbers, etc.

The parallel presence of ensembles of the binary sub-alphabets at different fields of information physiology is marked: at genetics, at color perception in vision physiology, etc. For example, an ensemble of three binary oppositions exists in a color circle of visual perception; on the basis of this parallelism, the author offers to denote genetic triplets by color and to consider color variant of bi-periodic table of triplets [2, 6].

The special attention is given to linguistics, in which there is long ago a concept of binary oppositions and their ensembles as structural base of the different linguistic languages (N.S.Trubetskoi, R.O.Jacobson, Ch.J.L.Baily, F.Jacob, etc.). For a long time there is an opinion, that languages of human dialogue were formed not on an empty place, and they are continuation of genetic language or, anyway, are closely connected to it, confirming the idea of unification of information bases of organisms (for example see [7, 8]). The name of the monograph by Baily [8] "On the Yin and Yang nature of language" is characteristic in this view.

Makovskii marks in his "Linguistic genetics" [9]: "The opinion about language as about living organism, which submitted to the natural laws of a nature, ascends to a deep antiquity ... Research of a nature, of character and of reasons of isomorphism between genetic and linguistic laws is one of the most important cardinal problems for linguistics of our time".

In a view of large number of such physiologic and linguistic materials, the author put forward a hypothesis that ensembles of binary sub-alphabets play a role of unification (unitized) base not only for linguistic languages but also for all or for majority of biological languages.

References:

[1] Petoukhov, S.V. Genetic Code and the Ancient Chinese "Book of Changes". *Symmetry: Culture and Science, v. 10, № 3-4, p.211-226*, 1999.

[2] Petoukhov S.V. *The Bi-periodic Table of Genetic Code and Number of Protons.* /Foreword by K. V. Frolov. Moscow, 258 p., 2001 (in Russian).

[3] Wittmann, H.G. Ansatze zur Entschlusselung des genetishen Codes. – *Die Naturwissenschaften*, B.48, 24, S. 55, 1961.

[4] Wilhelm, H. *Change: Eight Lectures on the I Ching*, N.Y. , 1960.

[5] He M. Symmetry in Structure of Genetic Code. *Proceedings of the Third All-Russian Interdisciplinary Scientific Conference "Ethics and the Science of Future. Unity in Diversity"*, Moscow, February 12-14, 2003.

[6] Petoukhov, S. V. Binary sub-alphabets of genetic language and problem of unification bases of biological languages, *IX International Conference "Mathematics, computer, education"*, Russia, Dubna, January 28-31, 191, 2002 (in Russian).

[7] Jacob, F. Le modele linquistique en biologie, *Critique*, Mars, tome XXX, № 322, 306-315, 1974.

[8] Baily Ch.J.L. *On the Yin and Yang nature of language.* – Ann Arbor, 1982.

[9] Makovskii M.M. *Linguistic Genetics.* – Moscow, Nauka, 1992 (in Russian).