PROCEEDINGS OF
THE 2010 INTERNATIONAL CONFERENCE ON
BIOINFORMATICS & COMPUTATIONAL BIOLOGY

BIOCOMP 2010

Volume II

Editors

Hamid R. Arabnia
Quoc-Nam Tran

Associate Editors

Rui Chang, Matthew He
Andy Marsh, Ashu M. G. Solo
Jack Y. Yang

WORLDCOMP'10
July 12-15, 2010
Las Vegas Nevada, USA
www.world-academy-of-science.org

MATRIX GENETICS AND THE BIPOLAR ALGEBRA OF THE GENETIC CODE

M. X. He[1] and S.V. Petoukhov[2]

[1]Division of Math, Science and Technology
Nova Southeastern University
Ft. Lauderdale, FL 33314 USA
[2]Department of Biomechanics, Mechanical Engineering Research Institute
Russian Academy of Sciences
Moscow, 101830 Russia

Abstract - Analysis of genetic code elements by matrix methods were taken from the theory of digital communication and of noise immunity signals. In this paper we use matrix approach to show that the degeneracy of the vertebrate mitochondrial code is agreed with the 8-dimensional algebra, which is unknown in modern mathematical natural science. This algebra allows one to reveal hidden peculiarities of the structure and evolution of the genetic code. We propose a new algebraic system for investigations in bioinformatics and mathematical biology including new approaches for the problem of noise immunity and classification of genetic molecules.

Keywords: Matrix genetics, genetic code, bipolar algebra.

1. Introduction

The history of science knows ideas about a special mathematics of living matter. For example, V. Vernadskiy (1965) [16] has put forward a hypothesis about a non-Euclidean geometry which dictates structural specifics of living matter. But how can one search such a general geometry of biological organisms if biological forms are so diverse? The fact that all biological organisms share the same molecular bases of the genetic code has provided a great unification of biological objects. Owning to this fact one may suggest that a special mathematics of living matter can be solved if the science will understand mathematical bases of the genetic code. Recent advances in bioinformatics have even led to a new definition of life: "Life is a partnership between genes and mathematics" (Stewart, 1999) [14].

What kind of mathematics dose a genetic code consist with in partner relations and define the structure and properties of living matter? This paper presents data about a 8-dimensional matrix algebra which is a candidate for a role of such genetic mathematics (first of all, in questions of the degeneracy of the genetic code). A discovery of deep connections of the genetic code with a multidimensional numeric system and its matrix algebras is described. These data are obtained on the basis of matrix approach to ensembles of molecular elements of the genetic code.

2. The Kronecker Family of Matrices of the Genetic Code

The mathematical theory of discrete signals uses Kronecker families of Hadamard matrices (Ahmed, & Rao, 1975) [1]: $H_{n+1}=[1\ 1;\ -1\ 1]^{(n)}$, where (n) means the integer Kronecker power. By analogy we use the Kronecker family of matrix presentations of the genetic code:

$$P^{(n)} = [C\ A;\ U\ G]^{(n)}, \qquad (1)$$

where C, A, U/T, G are the letters of the genetic alphabet (cytosine, adenine, uracil/thymine, guanine). The third Kronecker power of the kernel alphabetical matrix P = [C A; U G] gives the matrix $P^{(3)}$ = [C A; U G]$^{(3)}$ of the 64 triplets in a certain succession:

$$P = \begin{vmatrix} C & A \\ U & G \end{vmatrix};\quad P^{(3)} =$$

CCC	CCA	CAC	CAA	ACC	ACA	AAC	AAA
CCU	CCG	CAU	CAG	ACU	ACG	AAU	AAG
CUC	CUA	CGC	CGA	AUC	AUA	AGC	AGA
CUU	CUG	CGU	CGG	AUU	AUG	AGU	AGG
UCC	UCA	UAC	UAA	GCC	GCA	GAC	GAA
UCU	UCG	UAU	UAG	GCU	GCG	GAU	GAG
UUC	UUA	UGC	UGA	GUC	GUA	GGC	GGA
UUU	UUG	UGU	UGG	GUU	GUG	GGU	GGG

Fig. 1. The first genetic matrices $P^{(1)}$ and $P^{(3)}$ =[C A; U G]$^{(3)}$ of the Kronecker family

The genetic matrix [C A; U G]$^{(3)}$ on Figure 1 contains 16 sub-quadrants (2x2). Each of these sub-quadrants contains a subfamily of 4 NN-triplets with the same letters on their first two positions. An example of such subfamily is the set of four triplets CAC, CAA, CAU, and CAG with the same two letters CA on their first two positions.

The modern genetics has many dialects of the genetic code (see NCBI's website http://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi). The vertebrate mitochondria genetic code is considered as the most ancient and the most symmetrical code among these dialects (Frank-Kamenetskiy, 1988) [3]. This code has 8 subfamilies of NN-triplets, the code meaning of which is determined only by two first positions in each of these triplets (see Figure 2) whose meaning does not depend on the third position. We name the triplets, which belong to such "two-position" subfamilies of NN-triplets, as "black" triplets). This code has other 8 subfamilies of NN-triplets as well, the code meaning of which is determined by all its three positions whose meaning depends on the third position. We name the triplets, which belong to such "three-position" subfamilies of NN-triplets, as "white" triplets). A mosaic of a disposition of the black and white triplets in the genetic matrix [C A; U G]$^{(3)}$ reflects the specificity of the degeneracy of this basic dialect of the genetic code. It is unexpected phenomenological fact that these black and white triplets are disposed symmetrically in the matrix [C A; U G]$^{(3)}$ which is constructed formally without taking into account the data about the degeneracy of the genetic code (Figure 2). Really, the corresponding black-and-white mosaic (Figure 2) has the following symmetric features:

- The left and right halves of the matrix mosaic are mirror-anti-symmetric to each other in its colours: any pair of cells, disposed by mirror-symmetrical manner in these halves, possesses the opposite colours.
- The black-and-white matrix mosaic has a symmetric figure of a diagonal cross: diagonal quadrants of the matrix are equivalent to each other from the viewpoint of their mosaic.
- The genomatrix [C A; U G]$^{(3)}$ consists of the four pairs of neighbour rows with even and odd numeration numbers in each pair: 0-1, 2-3, 4-5, 6-7. The rows of each pair are equivalent to each other from the viewpoint of a disposition of the same amino acids in their appropriate cells.
- Mosaics of all rows have a meander-line character, which is connected with Rademacher functions from the theory of discrete signals processing.

CCC Pro	CCA Pro	CAC His	CAA Gln	ACC Thr	ACA Thr	AAC Asn	AAA Lys
CCU Pro	CCG Pro	CAU His	CAG Gln	ACU Thr	ACG Thr	AAU Asn	AAG Lys
CUC Leu	CUA Leu	CGC Arg	CGA Arg	AUC Ile	AUA Met	AGC Ser	AGA Stop
CUU Leu	CUG Leu	CGU Arg	CGG Arg	AUU Ile	AUG Met	AGU Ser	AGG Stop
UCC Ser	UCA Ser	UAC Tyr	UAA Stop	GCC Ala	GCA Ala	GAC Asp	GAA Glu
UCU Ser	UCG Ser	UAU Tyr	UAG Stop	GCU Ala	GCG Ala	GAU Asp	GAG Glu
UUC Phe	UUA Leu	UGC Cys	UGA Trp	GUC Val	GUA Val	GGC Gly	GGA Gly
UUU Phe	UUG Leu	UGU Cys	UGG Trp	GUU Val	GUG Val	GGU Gly	GGG Gly

Fig. 2. The representation of the genomatrix [C A; U G]$^{(3)}$ for the case of the vertebrate mitochondrial genetic code. The matrix contains 64 triplets and 20 amino acids with their traditional abbreviations. Stop-codons are marked as "Stop".

This symmetrical character of the degeneracy of the genetic code, which is presented by the matrix mosaic, is the key for many secrets of the genetic code. Let us investigate two initial questions: 1) what kind of mathematics has partnership relations with such mosaic matrix of the genetic code? In other words, whether is it possible to find the substantial mathematical justification to such a choice of nature? 2) Whether is this character of the degeneracy of the genetic code accidental (F. Crick (1968) [2] has stated a hypothesis of "the frozen case", i.e. an accidental character of this degeneracy)?

This alphabetic algorithm of digitization of 64 triplets is based on utilizing the two following binary-oppositional attributes of the genetic letters A, C, G, U/T: "purine or pyrimidine" and "2 or 3" hydrogen bonds. It uses also the famous thesis of molecular genetics that different positions inside triplets have different code meanings (Konopelchenko, & Rumer, 1975) [9]. In view of this "alphabetic" algorithm, the transformation of the genomatrix [C A; U G]$^{(3)}$ into the matrix YY_8 (Figure 3) is not an abstract and arbitrary action at all, but such a transformation can be utilized by bio-computer systems of organisms materially. By this alphabetic algorithm each triplet is read in the following way:

- Two first positions of each triplet are filled out by the symbol "α" instead of the complementary letters C and G on these positions and by the symbol "β" instead of the complementary letters A and U correspondingly;

- The third position of each triplet is filled out by the symbol "γ" instead of the pyrimidine (C or U) on this position and by the symbol "δ" instead of the purine (A or G) correspondingly;
- The triplets, which have the letters C or G in their first position, receive the sign "-" in those cases only for which their second position is occupied by the letter A. The triplets, which have the letters A or U on their first position, receive the sign "+" in those cases only for which their second positions is occupied by the letter C.

For example, the triplet CAG receives the symbol "-αβδ", because its first letter C is symbolized by "α", its second letter A is symbolized by "β", and its third letter G is symbolized by "δ". This triplet possesses the sign "-" because its first position has the letter C and its second position has the letter A. One can see that this algorithm recodes all triplets from the traditional alphabet C, A, U, G into the new alphabet α, β, γ, δ. In the result, each triplet receives one of the following 8 expressions: $ααγ = x_0$, $ααδ = x_1$, $αβγ = x_2$, $αβδ = x_3$, $βαγ = x_4$, $βαδ = x_5$, $ββγ = x_6$, $ββδ = x_7$ (Figure 3). We will suppose that the symbols "α", "β", "γ", "δ" are real numbers. This algorithm transforms the initial symbolic matrix [C A; U G]$^{(3)}$ into the numeric matrix YY_8 with the 8 coordinates x_0, x_1, x_2, x_3, x_4, x_5, x_6, x_7. We shall name these matrix components x_0, x_1,..., x_7, which are real numbers, as the "*YY*-coordinates".

0	1	2	3	4	5	6	7
CCC ααγ x_0	CCA ααδ x_1	CAC -αβγ $-x_2$	CAA -αββ $-x_3$	ACC βαγ x_4	ACA βαδ x_5	AAC -ββγ $-x_6$	AAA -βββ $-x_7$
CCU ααγ x_0	CCG ααδ x_1	CAU -αβγ $-x_2$	CAG -αβδ $-x_3$	ACU βαγ x_4	ACG βαδ x_5	AAU -ββγ $-x_6$	AAG -ββδ $-x_7$
CUC αβγ x_2	CUA αβδ x_3	CGC ααγ x_0	CGA ααδ x_1	AUC -ββγ $-x_6$	AUA -ββδ $-x_7$	AGC -βαγ $-x_4$	AGA -βαδ $-x_5$
CUU αβγ x_2	CUG αβδ x_3	CGU ααγ x_0	CGG ααδ x_1	AUU -ββγ $-x_6$	AUG -ββδ $-x_7$	AGU -βαγ $-x_4$	AGG -βαδ $-x_5$
UCC βαγ x_4	UCA βαδ x_5	UAC -ββγ $-x_6$	UAA -βββ $-x_7$	GCC ααγ x_0	GCA ααδ x_1	GAC -αβγ $-x_2$	GAA -αβδ $-x_3$
UCU βαγ x_4	UCG βαδ x_5	UAU -ββγ $-x_6$	UAG -βββ $-x_7$	GCU ααγ x_0	GCG ααδ x_1	GAU -αβγ $-x_2$	GAG -αβδ $-x_3$
UUC -ββγ $-x_6$	UUA -ββδ $-x_7$	UGC -βαγ $-x_4$	UGA -βαδ $-x_5$	GUC αβγ x_2	GUA αβδ x_3	GGC ααγ x_0	GGA ααδ x_1
UUU -ββγ $-x_6$	UUG -ββδ $-x_7$	UGU -βαγ $-x_4$	UGG -βαδ $-x_5$	GUU αβγ x_2	GUG αβδ x_3	GGU ααγ x_0	GGG ααδ x_1

$YY_8 =$

0	1	2	3	4	5	6	7
x_0	x_1	$-x_2$	$-x_3$	x_4	x_5	$-x_6$	$-x_7$
x_2	x_3	x_0	x_1	$-x_6$	$-x_7$	$-x_4$	$-x_5$
x_4	x_5	$-x_6$	$-x_7$	x_0	x_1	$-x_2$	$-x_3$
x_6	x_7	$-x_4$	$-x_5$	x_2	x_3	x_0	x_1

Fig. 3. The result of the algorithmic transformation of 64 triplets in the genomatrix [C A; U G]$^{(3)}$ into the numeric coordinates x_0, x_1, ..., x_7, which are based on the four symbols "α", "β", "γ", "δ". The bottom matrix YY_8 has only the numeric coordinates from the upper matrix. Its black (white) cells contain positive (negative) values of the YY-coordinates. Numeration of columns is shown.

3. The Analysis of the Algebraic Properties of the Matrix YY_8

A decomposition of the 8-parametric matrix YY_8 (Figure 3) leads to its presentation as a sum of the 8 basic matrices, each of which is connected with one of the coordinates x_0, x_1, x_2, x_3, x_4, x_5, x_6, x_7. Let us symbolize any basic matrix, which is related to any of *YY*-coordinates x_0, x_2, x_4, x_6 with even indexes, by the symbol $\mathbf{f}_k$ (where "f" is the first letter of the word "female" and $k = 0, 2, 4, 6$). And let us symbolize any matrix, which is related to any of *YY*-coordinates x_1, x_3, x_5, x_7 with odd indexes, by the symbol $\mathbf{m}_s$ (where "m" is the first letter of the word "male" and $s = 1, 3, 5, 7$). In this case one can present the matrix YY_8 by the expression (2),

$$YY_8 = x_0*\mathbf{f}_0+x_1*\mathbf{m}_1+x_2*\mathbf{f}_2+x_3*\mathbf{m}_3+x_4*\mathbf{f}_4+x_5*\mathbf{m}_5+x_6*\mathbf{f}_6+x_7*\mathbf{m}_7 \quad (2)$$

The important and unexpected fact is that the set of these 8 basic matrices $\mathbf{f}_0$, $\mathbf{m}_1$, $\mathbf{f}_2$, $\mathbf{m}_3$, $\mathbf{f}_4$, $\mathbf{m}_5$, $\mathbf{f}_6$, $\mathbf{m}_7$ forms the closed set relative to multiplications: a multiplication between any two matrices from this set generates a matrix from this set again. The table on Figure 4 presents the results of multiplications among these 8 matrices. The result of multiplying any two basic elements, which are taken from the left column and the upper row, is shown in the cell on the intersection of its row and column (for example, in accordance with this multiplication table $\mathbf{f}_2*\mathbf{m}_5 = -\mathbf{m}_7$).

	$\mathbf{f}_0$	$\mathbf{m}_1$	$\mathbf{f}_2$	$\mathbf{m}_3$	$\mathbf{f}_4$	$\mathbf{m}_5$	$\mathbf{f}_6$	$\mathbf{m}_7$
$\mathbf{f}_0$	$\mathbf{f}_0$	$\mathbf{m}_1$	$\mathbf{f}_2$	$\mathbf{m}_3$	$\mathbf{f}_4$	$\mathbf{m}_5$	$\mathbf{f}_6$	$\mathbf{m}_7$
$\mathbf{m}_1$	$\mathbf{f}_0$	$\mathbf{m}_1$	$\mathbf{f}_2$	$\mathbf{m}_3$	$\mathbf{f}_4$	$\mathbf{m}_5$	$\mathbf{f}_6$	$\mathbf{m}_7$
$\mathbf{f}_2$	$\mathbf{f}_2$	$\mathbf{m}_3$	$-\mathbf{f}_0$	$-\mathbf{m}_1$	$-\mathbf{f}_6$	$-\mathbf{m}_7$	$\mathbf{f}_4$	$\mathbf{m}_5$
$\mathbf{m}_3$	$\mathbf{f}_2$	$\mathbf{m}_3$	$-\mathbf{f}_0$	$-\mathbf{m}_1$	$-\mathbf{f}_6$	$-\mathbf{m}_7$	$\mathbf{f}_4$	$\mathbf{m}_5$
$\mathbf{f}_4$	$\mathbf{f}_4$	$\mathbf{m}_5$	$\mathbf{f}_6$	$\mathbf{m}_7$	$\mathbf{f}_0$	$\mathbf{m}_1$	$\mathbf{f}_2$	$\mathbf{m}_3$
$\mathbf{m}_5$	$\mathbf{f}_4$	$\mathbf{m}_5$	$\mathbf{f}_6$	$\mathbf{m}_7$	$\mathbf{f}_0$	$\mathbf{m}_1$	$\mathbf{f}_2$	$\mathbf{m}_3$
$\mathbf{f}_6$	$\mathbf{f}_6$	$\mathbf{m}_7$	$-\mathbf{f}_4$	$-\mathbf{m}_5$	$-\mathbf{f}_2$	$-\mathbf{m}_3$	$\mathbf{f}_0$	$\mathbf{m}_1$
$\mathbf{m}_7$	$\mathbf{f}_6$	$\mathbf{m}_7$	$-\mathbf{f}_4$	$-\mathbf{m}_5$	$-\mathbf{f}_2$	$-\mathbf{m}_3$	$\mathbf{f}_0$	$\mathbf{m}_1$

Fig. 4. The multiplication table of the basic matrices $\mathbf{f}_0$, $\mathbf{m}_1$, $\mathbf{f}_2$, $\mathbf{m}_3$, $\mathbf{f}_4$, $\mathbf{m}_5$, $\mathbf{f}_6$, $\mathbf{m}_7$ of the matrix YY_8 from Figures 3 and equatoin (2).

The multiplication table on Figure 4 defines the genetic 8-dimensional algebra YY_8. Multiplication of any two members of the octet algebra YY_8 generates a new member of the same algebra. This situation is similar to the situation of real numbers (or of complex numbers, or of hypercomplex numbers) when multiplication of any two members of the numeric system generates a new member of the same numerical system. In other words, the expression $YY_8=x_0*\mathbf{f}_0+x_1*\mathbf{m}_1+x_2*\mathbf{f}_2+x_3*\mathbf{m}_3+x_4*\mathbf{f}_4+x_5*\mathbf{m}_5+x_6*\mathbf{f}_6+x_7*\mathbf{m}_7$ is some kind of 8-dimensional numbers ("octet genonumber"). We mark this algebra and these octet genonumbers by the same symbol YY_8 accordingly.

One should pay special attention to the cells on the main diagonal of the multiplication table (Figure 4). These cells contain squares of the basic elements. In cases of hypercomplex numbers these diagonal cells contain elements "±1" typically. In our case these diagonal cells contain no real units at all but all diagonal cells are occupied by elements "$\pm\mathbf{f}_0$" and "$\pm\mathbf{m}_1$". Thereby the set of the 8 basic matrices $\mathbf{f}_0$, $\mathbf{m}_1$, $\mathbf{f}_2$, $\mathbf{m}_3$, $\mathbf{f}_4$, $\mathbf{m}_5$, $\mathbf{f}_6$, $\mathbf{m}_7$ is divided into two equal subsets by criterion of their squares. The first subset consists of elements with the even indexes: $\mathbf{f}_0$, $\mathbf{f}_2$, $\mathbf{f}_4$, $\mathbf{f}_6$. The squares of members of this $\mathbf{f}_0$-subset are equal to $\pm\mathbf{f}_0$ always. The second subset consists of elements with the odd indexes: $\mathbf{m}_1$, $\mathbf{m}_3$, $\mathbf{m}_5$, $\mathbf{m}_7$. The squares of members of this $\mathbf{m}_1$-subset are equal to $\pm\mathbf{m}_1$ always.

The basic element $\mathbf{f}_0$ possesses all properties of the real unit in relation to the members of the $\mathbf{f}_0$-subset: $\mathbf{f}_0^2=\mathbf{f}_0$, $\mathbf{f}_0*\mathbf{f}_2=\mathbf{f}_2*\mathbf{f}_0=\mathbf{f}_2$, $\mathbf{f}_0*\mathbf{f}_4=\mathbf{f}_4*\mathbf{f}_0=\mathbf{f}_4$, $\mathbf{f}_0*\mathbf{f}_6=\mathbf{f}_6*\mathbf{f}_0=\mathbf{f}_6$. But the element $\mathbf{f}_0$ does not possess the commutative property of real unit in relation to the members of the $\mathbf{m}_1$-subset: $\mathbf{f}_0*\mathbf{m}_p \neq \mathbf{m}_p*\mathbf{f}_0$, where p = 1,3,5,7. For this reason $\mathbf{f}_0$ is named "quasi-real unit from the $\mathbf{f}_0$-subset".

The basic element $\mathbf{m}_1$ possesses all properties of the real unit in relation to the members of the $\mathbf{m}_1$-subset: $\mathbf{m}_1^2=\mathbf{m}_1$, $\mathbf{m}_1*\mathbf{m}_3=\mathbf{m}_3*\mathbf{m}_1=\mathbf{m}_3$, $\mathbf{m}_1*\mathbf{m}_5=\mathbf{m}_5*\mathbf{m}_1=\mathbf{m}_5$, $\mathbf{m}_1*\mathbf{m}_7=\mathbf{m}_7*\mathbf{m}_1=\mathbf{m}_7$. But the element $\mathbf{m}_1$ does not possess the commutative property of real unit in relation to the members of the $\mathbf{f}_0$-subset: $\mathbf{m}_1*\mathbf{f}_k \neq \mathbf{f}_k*\mathbf{m}_1$, where k = 0,2,4,6. For this reason $\mathbf{m}_1$ is named "quasi-real unit from the $\mathbf{m}_1$-subset".

The principle "even-odd" exists in this algebra YY_8. Really all members of the $\mathbf{f}_0$-subset and their coordinates x_0, x_2, x_4, x_6 have even indexes and they are disposed in columns with the even numbers 0, 2, 4, 6 in the matrix YY_8 and in its multiplication table as well. All members of the $\mathbf{m}_1$-subset and their coordinates x_1, x_3, x_5, x_7 have the odd indexes and they are disposed in columns with the odd numbers 1, 3, 5, 7 in the matrix YY_8 (Figure 3) and in its multiplication table (Figure 4) as well. In accordance with Pythagorean and Ancient-Chinese traditions, all even numbers are named "female" numbers or Yin-numbers, and all odd numbers are named "male" numbers or Yang-numbers. From the viewpoint of this tradition, the elements $\mathbf{f}_0$, $\mathbf{f}_2$, $\mathbf{f}_4$, $\mathbf{f}_6$, x_0, x_2, x_4, x_6 with the even indexes play the role of "female" elements or Yin-elements, and the elements $\mathbf{m}_1$, $\mathbf{m}_3$, $\mathbf{m}_5$, $\mathbf{m}_7$, x_1, x_3, x_5, x_7 with the odd indexes play the role of "male" or Yang-elements. We name this algebra YY_8 as bipolar algebra. But it can be named also as the octet Yin-Yang-algebra, or the even-odd-algebra, or the bisex-algebra. Such algebra, which possesses two quasi-real units and no one real unit, gives new effective possibilities to model binary oppositions in biological objects at different levels, including sets of triplets, amino acids, male and female gametal cells, male and female chromosomes, etc.

In comparison with hypercomplex numbers, which have the real unit in the set of their basic elements, bipolar numbers YY_8 are the new category of numbers in the mathematical natural sciences in principle. In our view, knowledge of this category of numbers is necessary for deep understanding of biological phenomena, and, perhaps, it will be useful for mathematical natural sciences in the whole. Mathematical theory of bipolar numbers gives new formal and conceptual apparatus to model phenomena of reproduction and self-organization in living nature.

It can be demonstrated easily that bipolar algebras are the special generalization of the algebras of hypercomplex numbers in the form of "double-hypercomplex" numbers. Bipolar numbers become the appropriate hypercomplex numbers in those cases when all their female (or male) coordinates are equal to zero. Traditional hypercomplex numbers can be represented as a "mono-polar" half of appropriate bipolar numbers. The algorithm of such generalization is described in (Petoukhov, 2008a,b; Petoukhov & He, 2009) [12, 13, 14]. We denote bipolar numbers by double letters (for example, YY)

to distinguish them from traditional (complex and hypercomplex) numbers.

If all male coordinates are equal to 0 ($x_1 = x_3 = x_5 = x_7 = 0$), the bipolar numbers YY_8 become the Yin-genoquaternions $G_f = x_0*\mathbf{f_0} + x_2*\mathbf{f_2} + x_4*\mathbf{f_4} + x_6*\mathbf{f_6}$. If all female coordinates are equal to 0 ($x_0 = x_2 = x_4 = x_6 = 0$), the numbers YY_8 become the Yang-genoquaternions $G_m = x_1*\mathbf{m_1} + x_3*\mathbf{m_3} + x_5*\mathbf{m_5} + + x_7*\mathbf{m_7}$. These genetic quaternions G_f and G_m have the identical multiplication tables, which differ from the multiplication table of Hamilton quaternions (Petoukhov, 2008a,b; Petoukhov & He, 2009) [12, 13, 14]. Each of these genoquaternions corresponds to a case of an anisotropic space that provokes heuristic associations with anisotropic features of biological phenomena. Taking these facts into account, the octet genonumbers YY_8 can be named "the double genetic quaternions". It causes heuristic associations with a double helix of DNA, which is the bearer of genetic information. Just as the structure of three-dimensional physical space corresponds to the algebra of quaternions by Hamilton, so the structure of the genetic code corresponds to the algebra of the double genoquaternions.

4. The Structural Analogies Between the Genomatrix and the Bipolar Matrix YY_8

The main interest of bioinformatics to the octet bipolar algebra is connected with a possibility of its use as an adequate model of the structure of the genetic code. This possibility depends on structural coincidences between the bipolar matrix YY_8 and the genetic matrix [C A; G U]$^{(3)}$. A list of such non-trivial coincidences includes the following ones:

1. The first coincidence: The black-and-white mosaics of the bipolar matrix YY_8 and the genetic matrix [C A; G U]$^{(3)}$ are identical. (By an unknown reason, nature has divided the set of the 64 genetic triplets into two subset of 32 black triplets and 32 white triplets, which are disposed in the cells of 32 positive coordinates and 32 negative coordinates of the bipolar matrix YY_8).
2. The second coincidence: In the bipolar matrix YY_8, the pairs of the adjacent rows 0-1, 2-3, 4-5, 6-7 are identical to each other by the assortment and the disposition of numeric coordinates $x_0, x_1, x_2, x_3, x_4, x_5, x_6, x_7$. In the genetic matrix [C A; G U]$^{(3)}$, the same pairs of adjacent rows 0-1, 2-3, 4-5, 6-7 are identical each to each other by the assortment and the disposition of amino acids and stop-codons.
3. The third coincidence: In the bipolar matrix YY_8, the female coordinates x_0, x_2, x_4, x_6 occupy the columns with the even numbers 0, 2, 4, 6, and the male coordinates x_1, x_3, x_5, x_7 occupy the columns with the odd numbers 1, 3, 5, 7. In the genetic matrix [C A; G U]$^{(3)}$, the triplets with pyrimidine C or U on their third positions occupy the columns with the even numbers 0, 2, 4, 6; and the triplets with purine A or G on their third positions occupy the columns with the odd numbers 1, 3, 5, 7.
4. The fourth coincidence: In the bipolar matrix YY_8, one half of the quantity of the numeric coordinates (x_0, x_1, x_2, x_3) exists in the two quadrants along the main diagonal only; the second half of the numeric coordinates (x_4, x_5, x_6, x_7) exists in the two quadrants along the second diagonal only. In the genetic matrix [C A; G U]$^{(3)}$, one half of kinds of amino acids exists in the two quadrants along the main diagonal only (Ala, Arg, Asp, Gln, Glu, Gly, His, Leu, Pro, Val); the second half of kinds of amino acids exists in the two quadrants along the second diagonal only (Asn, Cys, Ile, Lys, Met, Phe, Ser,Thr, Trp, Tyr).
5. The fifth coincidence: In the bipolar matrix YY_8, those six kinds of different numeric matrices are generated by means of some kinds of permutations of columns and rows of this matrix, each of which possesses its own kind of the 8-dimensional bipolar algebra.

In the genetic matrix [C A; G U]$^{(3)}$, the same six kinds of permutations of columns and rows fit the six possible kinds of permutations of positions inside the 64 triplets (1-2-3, 2-3-1, 3-1-2, 3-2-1, 2-1-3, 1-3-2), which lead to the new genomatrices with symmetric and interrelated mosaics (this fifth coincidence is explained additionally in the works (Petoukhov, 2008; Petoukhov & He, 2009).

One should note that the black cells of the genomatrix [C A; U G]$^{(3)}$ contain the black *NN*-triplets, which encode the 8 high-degeneracy amino acids(Ala, Arg, Gly, Leu, Pro, Ser, Thr, Val), each of which is encoded by 4 triplets or more in the considered basic dialect of the genetic code. The white cells of this genomatrix contain the white NN-triplets, which encode the 12 low-degeneracy amino acids (Asn, Asp, Cys, Gln, Glu, His, Ile, Lys, Met, Phe, Trp, Tyr), each of which is encoded by 3 triplets or less correspondingly.

The described structural coincidences of two matrices YY_8 and [C A; U G]$^{(3)}$ allow one to consider the octet bipolar algebra YY_8 as the adequate model of the structure of the genetic code. One can postulate such an algebraic model and then deduce some peculiarities of the genetic code from this model.

These results of the comparison analysis give the following answer to the question of mysterious principles of the degeneracy of the vertebrate mitochondrial genetic code from the viewpoint of the proposed algebraic model. The matrix disposition of the 20 amino acids and the stop-signals is determined by algebraic principles of the matrix disposition of the YY-coordinates. Moreover the disposition of the 32 black triplets and the high-degeneracy amino acids in this basic dialect of the genetic code is determined by the disposition of the YY-coordinates with the sign "+". And the disposition of the 32 white triplets, the low-degeneracy amino acids and stop-signals is determined by the disposition of the YY-coordinates with the sign "-". One can recall here that the division of the set of 20 amino acids into the two sub-sets of the 8 high-degeneracy amino acids and the 12 low-degeneracy amino acids is the invariant rule of all the dialects of the genetic code practically (Petoukhov, 2005). The described structural coincidences between both matrices do not exhaust the interconnections between the genetic code systems and the bipolar matrices.

5. Matrix Genetics and Applications of the Genetic Bipolar Algebra

The genetic systems from the viewpoint of matrix analysis are united under the general name "matrix genetics". This new scientific field is developed intensively and it has interesting results already in works (He, 2001, 2003a, 2003b; He, Petoukhov, 2007; He, Petoukhov, Ricci, 2004; Petoukhov, 2001, 2005, 2008a, 2008b; Petoukhov, He, 2009; etc.) [4, 5, 6, 7, 8, 10, 11, 12, 13, 14].

The discovery of the genetic bipolar algebra gives new possibilities for structural analysis of genetic systems. For example this matrix algebra allows revealing of internal structure in the set of 20 amino acids: this set appears to be structured by means of splitting into subsets of "female" amino acids, "male" amino acids and "androgynous" amino acids from the viewpoint of the bipolar matrix YY_8 (Fig. 2). Amino acid is named accordingly as a female amino acid if it corresponds only to YY-coordinates with even indexes x_0, x_2, x_4, x_6 (that are Yin-coordinates) in the bipolar matrix YY_8. One can see from Figure 2 that female amino acids are Asn, Asp, Cys, His, Ile, Phe, Tyr. Amino acid is named accordingly as a male amino acid if it corresponds only to YY-coordinates with odd indexes x_1, x_3, x_5, x_7 (that are Yang-coordinates) in the bipolar matrix YY_8. The male amino acids are Glu, Lys, Met, Gln, Trp. Amino acids is named as an androgynous if it corresponds simultaneously to YY-coordinates with even indexes and with odd indexes. The androgynous amino acids are the 8 high-degeneracy acids Ala, Arg, Gly, Pro, Thr, Val, Ser, Leu. The knowledge of internal structure of the set of 20 amino acids is useful for analysis of proteins structures and evolution of dialects of the genetic code (Petoukhov, 2008a; Petoukhov, He, 2009) [12, 14]. For example, all kinds of proteins can be classified as female, male of androgynous proteins depend on the fact what kind of amino acids (female, male or androgynous) dominates in a protein. Specificity of interactions between proteins of various sexual types is a new interesting problem for investigations.

The analysis of evolution of dialects of the genetic code from the viewpoint of the genetic bipolar algebra demonstrates that this evolution is a struggle between male and female beginnings, which reminds of the social struggle between male and female sexes (matriarchy and patriarchy, etc.). In a course of evolution, the "female" amino acids begin to occupy cells with the "male" amino acids in the genomatrix $[C\ A;\ U\ G]^{(3)}$ (Figure 2) but male triplets gain revenge in termination functions: they not only encode all stop-signals in all dialects, but they withdraw the start-codon function from female triplets in some extent. One can think that two – male and female – sexes of biological organisms have arisen not from a "white space" but they have predecessors at the molecular-genetic level (Petoukhov & He, 2009) [14]. Plato had formulated the famous statement about a congenital aspiration of each person to look for the second half. From the viewpoint of our "bisex" or bipolar conception, which is based on the genetic bipolar algebra, Plato's statement can be transferred into the world of those congenital properties of genetic molecules which are reflected in their search of their second halves.

One of interesting examples of such searching is given by histones. In eukaryote cells, filaments of DNA are coiled around nucleosomes, each of which is a shank consisting of the histones of the four types: H2A, H2B, H3 and H4. This set of four types is divided by nature into the pairs of one-specific histones. The histones H2A and H2B possess the important possibility to create the pair just one with another on the basis of their mutual revealing and mutual "attraction" in a molecular bouillon (by analogy with a male and a female individuals of one species among macroscopic biological organisms). Another pair consists of the histones H3 and H4, which possess the similar possibility to create the pairs just one with another on the analogical basis of

their mutual revealing and mutual "attraction" in molecular bouillon. Each nucleosome is formed in accordance with the principle of the multi-level recognition defined by the structures of the histones. In the first step, the spiral domains cooperate among themselves. As a result, pairs (dimers) arise: one pair H3-H4 and two pairs H2A-H2B. In the second step, two first dimers form the pair association of the following level of complexity: the tetramer arises with two pairs H3-H4. In the third step, this tetramer forms a pair association of the higher level with two pairs H2A-H2B. As a result, the octamer of the histones arises (http://www.ncbi.nlm.nih.gov). All these searches and copulations of one-specific histones into pairs, and then into new pairs from previous pairs occur in a molecular bouillon with a huge bedlam of biological molecules of other kinds and their splinters. It occurs despite of effects of electric shielding and other noise circumstances there. In our view, taking into account the described facts, one can put forward the working hypothesis about existence of "a sexual intermolecular attraction" (or a "bipolar attraction") between genetic one-specific elements as a new biophysical factor of a quantum mechanical sense. This new hypothetical factor or principle is presented, first of all, as an explanation of molecular-genetic phenomena of search the one-specific pair partner by multi-atomic bio-molecules to create a specific pair in complex conditions of multi-component bullion. The genetic bipolar algebra can be useful to model and investigate such a factor. This factor can have a force character and/or information character. It does not reject the existence of other known factors (for example, interactions of electric charges and so forth), but it is additional to them. Of course, it would be wrong to extend an action of this factor of "a sexual intermolecular attraction", which is proposed in connection with phenomena of assembly of pairs of one-specific multi-atomic molecular elements (multi-atomic quantum mechanical "modules"), into the field of all aspects of molecular-genetic organization.

References

[1] Ahmed, N., & Rao, K. (1975) *Orthogonal transforms for digital signal processing*. New-York: Springer-Verlag Inc.

[2] Crick, F. (1968). The origin of the genetic code. *J. of Molecular Biology, 38*, 367-379.

[3] Frank-Kamenetskiy, M. D. (1988). *The most principal molecule.* Moscow: Nauka (in Russian).

[4] He, M. (2001). Double helical sequences and doubly stochastic matrices. In S. Petoukhov (Ed.). *Symmetry in genetic information*, (pp. 307-330). Budapest: International Symmetry foundation.

[5] He, M. (2003a). Symmetry in Structure of Genetic Code. Proceedings of the 3rd All-Russian Interdisciplinary Scientific Conference "*Ethics and the Science of Future. Unity in Diversity*", Feb. 12-14, Moscow, pp. 80-85.

[6] He, M. (2003b). Genetic Code, Attributive Mappings and Stochastic Matrices, *Bulletin for Mathematical Biology*, Vol. 66/5, pp 965-973.

[7] He, M., and Petoukhov, S. (2007). Harmony of living nature, symmetries of genetic systems and matrix genetics. *International journal of integrative biology, 1*(1), pp.41-43.

[8] He, M., Petoukhov, S.V., & Ricci, P.E.(2004). Genetic code, Hamming distance and stochastic matrices. *Bulletin for Mathematical Biology, 66*(5), pp 965-973.

[9] Konopelchenko, B. G., & Rumer, Yu. B. (1975). Classification of the codons in the genetic code. *Doklady AN SSSR, 223*(2), 145-153 (in Russian).

[10] Petoukhov, S.V. (2001). Genetic codes 1: binary sub-alphabets, bi-symmetric matrices and golden section; Genetic codes 2: numeric rules of degeneracy and the chronocyclic theory. *Symmetry: Culture and Science, 12* (1), 255-306.

[11] Petoukhov, S.V. (2005a). The rules of degeneracy and segregations in genetic codes. The chronocyclic conception and parallels with Mendel's laws. In M. He, G.Narasimhan, & S.Petoukhov (Ed.), *Advances in Bioinformatics and its Applications, Series in Mathematical Biology and Medicine, 8*, (pp. 512-532). Singapore:World Scientific.

[12] Petoukhov, S.V. (2008a). *Matrix genetics, algebras of the genetic code, noise-immunity.* Moscow: RCD (in Russian).

[13] Petoukhov, S.V. (2008b). Matrix genetics, part 2: the degeneracy of the genetic code and the octave algebra with two quasi-real units (the "Yin-Yang octave algebra"). 1-27. Retrieved March 08, 2008, from http://arXiv:0803.3330.

[14] Petoukhov, S. and He, M. (2009). *Symmetrical Analysis Techniques for Genetic Systems and Bioinformatic: Advanced Patterns and Applications.* Hershey: IGI Global.

[15] Stewart I., (1999). *Life's other secret*: The new mathematics of the living world. New-York: Penguin.

[16] Vernadsky, V.I. (1965). *Chemical structure of the Earth and its surrounding.* Moscow: Nauka.