the Genetic Coding System and Unitary Matrices

Sergey V. Petoukhov[1]

[1]-Mechanical Engineering Research Institute of Russian Academy of Sciences. Russia, 101990, Moscow, M. Kharitonievskiy pereulok, 4,
http://eng.imash.ru/, info@imash.ru

Comment: Some materials of this article were presented by the author in his keynote speech at the international conference "Artificial Intelligence, Medical Engineering, Education" (Moscow, Russia, 21-23 August 2017, http://www.ruscnconf.org/aimee2017/index.html) and in his lecture at the international workshop of the Competence Center for Algorithmic and Mathematical Methods in Biology, Biotechnology and Medicine (Mannheim, Germany, 27 November 2017, https://www.cammbio.hs-mannheim.de/kick-off-event.html). They were also presented by the author in his keynote speech at the international Belgrade Bioinformatics Conference 2018 (Belgrade, Serbia, 18-22 June, 2018, http://belbi.bg.ac.rs/).

Abstract. Information molecules of DNA and RNA should obey principles of quantum mechanics where unitary operators in form of unitary matrices have key meanings. Unitary matrices are the basis of calculations in quantum computers. This article presents some author's results, which show that matrix forms of the representation of structured systems of molecular-genetic alphabets can be considered as sets of sparse unitary matrices related with phenomenologic features of the degeneracy of the genetic code. These sparse unitary matrices have orthogonal systems of functions in their rows and columns. A complementarity exists among some unitary genetic matrices in relation each other. Decompositions of numeric genetic matrices into sets of sparse unitary matrices are connected with the logical operation of modulo-2 addition used in quantum computers as well. Tensor (or Kronecker) families of unitary genetic matrices with their fractal-like properties are also considered. Hidden reqularities of frequencies of n-plets of hydrogen bonds are revealed in long DNA sequences and in complete sets of chromosomes of some tested organisms. The mathematical model of these reqularities was proposed on the basis of formalisms of quantum informatics. This model has some analogies with the Hardy-Weinberg law of population genetics. The possible representation of the set of the mentioned frequencies in forms of matrix representations of some 2^n-dimensional hypercomplex numbers was also noted.

Keywords: genetic code, alphabet, unitary matrix, dyadic shift, decomposition, spectral presentation, fractal, tensor product, hypercomplex numbers, quantum informatics

1. Introduction.

The information molecules of DNA and RNA of the genetic coding system belong to the world of molecules, in which the principles of quantum mechanics manage. This article presents results of the author's study of abilities of using formalisms of quantum mechanics and quantum informatics to model regular structures of molecular-genetic systems. First of all, we are talking about searching for correspondences between unitary operators and structured alphabets of DNA and RNA in their matrix forms of representation. The article provides additional materials for the development of quantum information modeling of structured molecular-genetic ensembles; elements of the quantum information modeling have been described in the author's work about tetra-group symmetries in long DNA texts [Petoukhov, 2017a].

    In line with one of the postulates of quantum mechanics, the evolution of a closed quantum system is described by unitary transformations. Computational processes in quantum computer science are based on unitary operators that serve as quantum gates. "*Any unitary matrix specifies a valid*

quantum gate" [Nielsen, Chuang, 2010, p. 18]. Any physical impact on a qubit in quantum mechanics is described by a linear unitary operator.

In quantum mechanics and quantum computer science, an important role is played by unitary operators in form of Hadamard matrices with complete orthogonal systems of Walsh functions in them. Hadamard operators are also widely used for spectral representations of signal vectors in the technique of noise-immune communication [Seberry, Wysocki, Wysocki, 2005], the sequency analysis of Harmuth [Harmuth, 1977, 1989], the digital logical holography [Derzhypolskyy, Melenevskyy, Gnatovskyy, 2007; Morita, Sakurai, 1973; Soroko, 1974] and algorithms of quantum informatics [Nielsen, Chuang, 2010]. But the Hadamard matrices with their complete systems of orthogonal functions are not the only unitary matrices with complete systems of orthogonal functions in them.

In this paper, we present other unitary matrices with other complete systems of orthogonal functions that were discovered by the author in the course of the algebraic modeling of molecular alphabets of DNA and RNA. These unitary matrices are sparse ones and they form sets of mutual-complementary matrices (in some algebraic sence). We conditionally call them unitary genetic matrices (or briefly, unitary geno-matrices). This makes us recall the well-known proposition that different natural systems may need - for their spectral analysis - in their own systems of orthogonal functions: "*after Fourier it was found that for some problems, harmonic sinusoids rather than other systems of orthogonal functions, for example, the Legendre polynomials, are better suited. In fact, any particular problem needs its own system of orthogonal functions. This was most clearly manifested in the course of the development of quantum mechanics*" [Soroko, 1973]. These unitary sparse geno-matrices contain complete systems of orthogonal functions and have special algebraic properties. They can serve as the basis for a new class of spectral representations of vectors in biology and other fields of science, as well as a new class of bio-mathematical models and algorithms in classical and quantum computer sciences.

One should add that quantum-information aspects of life are actively discussed in modern science, for example, in the book [Quantum aspects of life, 2008]; in articles about a biology of quantum information [Matsuno, 1999, 2003, 2015; Matsuno, Paton, 2000]; in articles about a possible meaning of the quantum algorithm of Grower in genetic information [Patel, 2001 a,b,c], etc.

2. Matrix representations of DNA-alphabets and genetic binary oppositions

Science does not know why the basic alphabet of DNA has been created by Nature from just four letters (adenine A, thymine T, cytosine C and guanine G), and why just these very simple molecules were chosen for the DNA-alphabet (out of millions of possible molecules). But science knows [Fimmel, Danielli, Strüngmann, 2013; Petoukhov, 2008; Petoukhov, He, 2009; Stambuk, 1999] that these four molecules are interrelated due to their symmetrical peculiarities into the united molecular ensemble with its three pairs of binary-oppositional traits or indicators (Fig. 1).

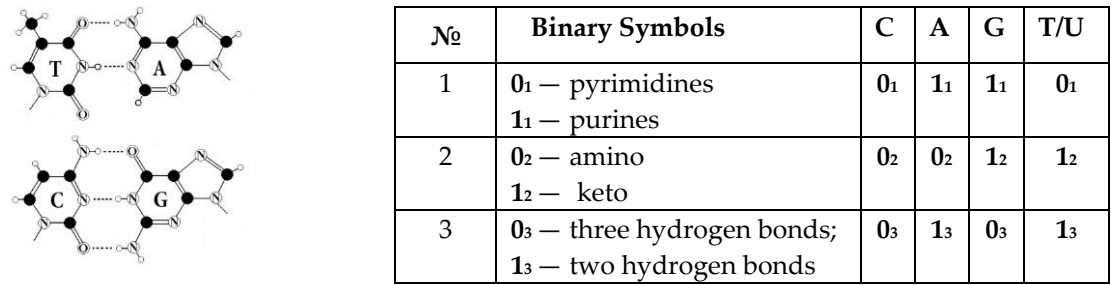| №	Binary Symbols	C	A	G	T/U
1	0_1 — pyrimidines 1_1 — purines	0_1	1_1	1_1	0_1
2	0_2 — amino 1_2 — keto	0_2	0_2	1_2	1_2
3	0_3 — three hydrogen bonds; 1_3 — two hydrogen bonds	0_3	1_3	0_3	1_3

Fig. 1. Left: the four nitrogenous bases of DNA: adenine A, guanine G, cytosine C, and thymine T. Right: three binary sub-alphabets of the genetic alphabet on the basis of three pairs of binary-oppositional traits or indicators.

These three pairs of binary-oppositional traits or indicators are the following:

(1) Two letters are purines (A and G), and the other two are pyrimidines (C and T). From the standpoint of these binary-oppositional traits one can denote C = T = 0, A = G = 1. From the standpoint of these

traits, any of the DNA-sequences are represented by a corresponding binary sequence. For example, GCATGAAGT is represented by 101011110;

(2) Two letters are amino-molecules (A and C) and the other two are keto-molecules (G and T). From the standpoint of these traits one can designate A = C = 0, G = T = 1. Correspondingly, the same sequence GCATGAAGT is represented by another binary sequence, 100110011;

(3) The pairs of complementary letters, A-T and C-G, are linked by 2 and 3 hydrogen bonds, respectively. From the standpoint of these binary traits, one can designate C = G = 0, A = T = 1. Correspondingly, the same sequence GCATGAAGT is read as 001101101.

Accordingly, each of the DNA-sequences of nucleotides is the carrier of three parallel messages on three different binary languages. At the same time, these three types of binary representations form a common logic set on the basis of the logic operation of modulo-2 addition denoted by the symbol $\oplus$: modulo-2 addition of any two such binary representations of the DNA-sequence gives the third binary representation of the same DNA-sequence: for example, $101011110 \oplus 100110011 = 001101101$. One can remind here the rules of the bitwise modulo-2 addition: $0 \oplus 0 = 0$; $0 \oplus 1 = 1$; $1 \oplus 0 = 1$; $1 \oplus 1 = 0$. The logic operation of modulo-2 addition is actively used in classical and quantum computers. Below we use the operation of modulo-2 addition for those decompositions of genetic matrices, which lead to interesting sets of inter-complementary unitary matrices of sparse types for special kinds of spectral representations of vectors.

Taking into account the phenomenological fact that each of DNA-letters C, A, T and G is uniquely defined by any two kinds of mentioned binary-oppositional indicators (Fig. 1), these genetic letters can be represented by means of corresponding pairs of binary symbols, for example, from the standpoint of two first binary-oppositional indicators. It is convenient for us - for the further description - use at the first position of each of letters its binary symbol from the second pair of binary-oppositional indicators (the indicator "amino or keto": C=A=0, T=G=1) and at the second positions of each of letters its binary symbol from the first pair of binary-oppositional indicators (the indicator "pyrimidine or purine": C=T=0, A=G=1). In this case the letter C is represented by the binary symbol $0_2 0_1$ (that is as 2-bit binary number), A – by the symbol $0_2 1_1$, T – by the symbol $1_2 0_1$, G – by the symbol $1_2 1_1$. Using these representations of separate letters, each of 16 doublets is represented as the concatenation of the binary symbols of its letters (that is as 4-bit binary number): for example, the doublet CC is represented as 4-bit binary number $0_2 0_1 0_2 0_1$, the doublet CA – as 4-bit binary number $0_2 0_1 0_2 1_1$, etc. By analogy, each of 64 triplets is represented as the concatenation of the binary symbols of its letters (that is as 6-bit binary number): for example, the triplet CCC is represented as 6-bit binary number $0_2 0_1 0_2 0_1 0_2 0_1$, the triplet CCA – as 6-bit binary number $0_2 0_1 0_2 0_1 0_2 1_1$, etc. In general, each of n-plets is represented as the concatenation of the binary symbols of its letters (below we will not show these indexes 2 and 1 of separate letters in binary representations of n-plets but will remember that each of positions corresponds to its own kind of indicators from the first or from the second set of indicators in Fig. 1).

It is convenient to represent DNA-alphabets of 4 nucleotides, 16 doublets, 64 triplets, … 4^n n-plets in a form of appropriate square tables (Fig. 2), which rows and columns are numerated by binary symbols in line with the following principle. Entries of each column are numerated by binary symbols in line with the first set of binary-oppositional indicators in Fig. 1 (for example, the triplet CAG and all other triplets in the same column are the combination "pyrimidine-purin-purin" and so this column is correspondingly numerated 011). By contrast, entries of each of rows are numerated by binary numbers in line with the second set of indicators (for example, the same triplet CAG and all other triplets in the same row are the combination "amino-amino-keto" and so this row is correspondingly numerated 001). In such tables (Fig. 2), each of 4 letters, 16 doublets, 64 triplets, … takes automatically its own individual place and all components of the alphabets are arranged in a strict order.

	0	1
0	C	A
1	T	G

	00	01	10	11
00	CC	CA	AC	AA
01	CT	CG	AT	AG
10	TC	TA	GC	GA
11	TT	TG	GT	GG

	000	001	010	011	100	101	110	111
000	CCC	CCA	CAC	CAA	ACC	ACA	AAC	AAA
001	CCT	CCG	CAT	CAG	ACT	ACG	AAT	AAG
010	CTC	CTA	CGC	CGA	ATC	ATA	AGC	AGA
011	CTT	CTG	CGT	CGG	ATT	ATG	AGT	AGG
100	TCC	TCA	TAC	TAA	GCC	GCA	GAC	GAA
101	TCT	TCG	TAT	TAG	GCT	GCG	GAT	GAG
110	TTC	TTA	TGC	TGA	GTC	GTA	GGC	GGA
111	TTT	TTG	TGT	TGG	GTT	GTG	GGT	GGG

Fig. 2. The square tables of DNA-alphabets of 4 nucleotides, 16 doublets and 64 triplets with a strict arrangement of all components. Each of tables is constructed in line with the principle of binary numeration of its column and rows on the basis of binary-oppositional traits of the nitrogenous bases (see explanations in the text).

It is essential that these 3 separate genetic tables form the joint tensor family of matrices since they are interrelated by the known operation of the tensor (or Kronecker) product of matrices (Fig. 3). So they are not simple tables but matrices. By definition, under tensor multiplication of two matrices, each of entries of the first matrix is multiplied with the whole second matrix [Bellman, 1960]. The second tensor power of the (2*2)-matrix [C, A; T, G] of 4 DNA-letters gives automatically the (4*4)-matrix of 16 doublets; the third tensor power of the same (2*2)-matrix of 4 DNA-letters gives the (8*8)-matrix of 64 triplets with the same strict arrangement of entries as in Fig. 2. In this tensor construction of the tensor family of genetic matrices, data about binary-oppositional traits of genetic letters C, A, T and G are not used at all. So, the structural organization of the system of DNA-alphabets is connected with the algebraic operation of the tensor product (Fig. 3). It is important since the operation of the tensor product is well known in mathematics, physics and informatics, where it gives a way of putting vector spaces together to form larger vector spaces. The following quotation speaks about the crucial meaning of the tensor product: «*This construction is crucial to understanding the quantum mechanics of multiparticle systems*» [Nielsen, Chuang, 2010, p. 71]. For us the most interesting is that the tensor product is one of basic instruments in quantum informatics.

$$\begin{bmatrix} C & A \\ T & G \end{bmatrix} \otimes \begin{bmatrix} C & A \\ T & G \end{bmatrix} = \begin{bmatrix} C*\begin{bmatrix} C & A \\ T & G \end{bmatrix} & A*\begin{bmatrix} C & A \\ T & G \end{bmatrix} \\ T*\begin{bmatrix} C & A \\ T & G \end{bmatrix} & G*\begin{bmatrix} C & A \\ T & G \end{bmatrix} \end{bmatrix} = \begin{bmatrix} CC & CA & AC & AA \\ CT & CG & AT & AG \\ TC & TA & GC & GA \\ TT & TG & GT & GG \end{bmatrix}$$

CCC	CCA	CAC	CAA	ACC	ACA	AAC	AAA
CCT	CCG	CAT	CAG	ACT	ACG	AAT	AAG
CTC	CTA	CGC	CGA	ATC	ATA	AGC	AGA
CTT	CTG	CGT	CGG	ATT	ATG	AGT	AGG
TCC	TCA	TAC	TAA	GCC	GCA	GAC	GAA
TCT	TCG	TAT	TAG	GCT	GCG	GAT	GAG
TTC	TTA	TGC	TGA	GTC	GTA	GGC	GGA
TTT	TTG	TGT	TGG	GTT	GTG	GGT	GGG

Fig. 3. The tensor family of genetic matrices [C, A; T, G]$^{(n)}$ (here tensor power n = 1, 2, 3) of DNA-alphabets of 4 nucleotides, 16 doublets and 64 triplets. The symbol ⊗ means the tensor product.

As is known, the degeneracy of the genetic code has the important specificity: the entire set of 64 triplets is divided by Nature into 2 equal binary-opposition subsets [Rumer, 1968]:

- 32 triplets with "strong roots" (black colors in Fig. 4), i.e., with 8 "strong" doublets AC, CC, CG, CT, GC, GG, GT, TC in their first positions;

- 32 triplets with "weak roots" (white colors in Fig. 4), i.e., with 8" weak" doublets CA, AA, AG, AT, GA, TA, TG, TT in their first positions.

	00	01	10	11
00	CC (00)	CA (01)	AC (10)	AA (11)
01	CT (01)	CG (00)	AT (11)	AG (10)
10	TC (10)	TA (11)	GC (00)	GA (01)
11	TT (11)	TG (10)	GT (01)	GG (00)

	000	001	010	011	100	101	110	111
000	CCC (000)	CCA (001)	CAC (010)	CAA (011)	ACC (100)	ACA (101)	AAC (110)	AAA (111)
001	CCT (001)	CCG (000)	CAT (011)	CAG (010)	ACT (101)	ACG (100)	AAT (111)	AAG (110)
010	CTC (010)	CTA (011)	CGC (000)	CGA (001)	ATC (110)	ATA (111)	AGC (100)	AGA (101)
011	CTT (011)	CTG (010)	CGT (001)	CGG (000)	ATT (111)	ATG (110)	AGT (101)	AGG (100)
100	TCC (100)	TCA (101)	TAC (110)	TAA (111)	GCC (000)	GCA (001)	GAC (010)	GAA (011)
101	TCT (101)	TCG (100)	TAT (111)	TAG (110)	GCT (001)	GCG (000)	GAT (011)	GAG (010)
110	TTC (110)	TTA (111)	TGC (100)	TGA (101)	GTC (010)	GTA (011)	GGC (000)	GGA (001)
111	TTT (111)	TTG (110)	TGT (101)	TGG (100)	GTT (011)	GTG (010)	GGT (001)	GGG (000)

Fig. 4. Black-and-white mosaics represent the distribution of strong and weak doublets in the matrix of 16 doublets (left) and the distribution of triplets with strong and weak roots in the matrix of 64 triplets (on the right). Binary number in brackets in each of matrix cells is a sum of modulo-2 addition of binary numberings of the row and the column of the cell.

Code meanings of triplets with strong roots do not depend on the letters on their third position; code meanings of triplets with weak roots depend on their third letter. What are locations of these

strong (black) and weak (white) members of DNA-alphabets in the genetic matrices shown in Figs. 2 and 3?

The unexpected phenomenological fact is a symmetrical location (Fig. 4) of all black and white entries in the genetic matrices of 16 doublets and 64 triplets, which were constructed very formally without any mention about amino acids and the degeneracy of the genetic code.

Symmetrical properties of mosaics in the genetic matrices in Fig. 4 are the following:
1) the left and right halves of the matrix mosaic are mirror-anti-symmetric each to other in its colors: any pair of cells, disposed by mirror-symmetrical manner in the halves, possesses the opposite colors;
2) the block mosaic of the matrix has the cruciform character: both quadrants along each diagonals are identical each other from the standpoint of their mosaic;
3) mosaic of each of rows has the meander character identical to known Rademacher functions
$r_n(t)$ = sign(sin$2^n\pi t$), n = 1, 2, 3,…, which are particular cases of Walsh functions and contain only values +1 and -1 (https://www.encyclopediaofmath.org/index.php/Rademacher_system).

Using this analogy with Rademacher-Walsh functions, one can represent the symbolic genetic matrices in Fig. 4 in forms of numeric matrices R_4 and R_8 with their entries +1 and -1 in Fig. 5 where numbers +1 (-1) represent black (white) doublets and triplets correspondingly. Taking into account that meander-like mosaics of rows of matrices R_4 and R_8 correspond Rademacher functions, we conditionally called these matrices "Rademacher matrices" in all our publications beginning from our book [Petoukhov, 2008] (although Hans Rademacher himself never worked with such matrices).

R_4 =					R_8 =							

R_4 =

1	-1	1	-1
1	1	-1	-1
1	-1	1	-1
-1	-1	1	1

R_8 =

1	1	-1	-1	1	1	-1	-1
1	1	-1	-1	1	1	-1	-1
1	1	1	1	-1	-1	-1	-1
1	1	1	1	-1	-1	-1	-1
1	1	-1	-1	1	1	-1	-1
1	1	-1	-1	1	1	-1	-1
-1	-1	-1	-1	1	1	1	1
-1	-1	-1	-1	1	1	1	1

Fig. 5. Numeric representations R_4 and R_8 of the genetic matrices of 16 doublets and 64 triplets from Fig. 4. Matrix cells with number +1 (-1) correspond cells with black (white) doublets and triplets in Fig. 4. Each of rows of the numeric matrices represents one of Rademacher-Walsh functions.

It should be noted that a huge quantity 64! ≈ 10^{89} of variants exists for dispositions of 64 triplets in a separate (8*8)-matrix. For comparison, the modern physics estimates time of existence of the Universe in 10^{17} seconds. It is obvious that an accidental disposition of black and white triplets in a separate (8*8)-matrix will give almost never any symmetry. But in our approach, this matrix of 64 triplets is not a separate matrix, but it is one of members of the family of matrices of genetic alphabets interrelated by means of binary-oppositional traits of nitrogenous bases A, T, C, G (and additionally it is one of members of the tensor family of matrices [C, A; T, G]$^{(n)}$ of interrelated alphabets of DNA).

These numeric matrices R_4 and R_8 with their mosaics (Fig. 5) represent the phenomenological peculiarities of the degeneracy of the genetic code. The exponentiation of these genetic matrices in the second power leads to their doubling and quadrupling: $R_4^2 = 2*R_4$ and $R_8^2 = 4*R_8$. This resembles the doubling and quadrupling the genetic material under mitosis and meiosis of biological cells. Let us analyze algebraic properties of these genetic matrices R_4 and R_8 more deeply.

3.   The genetic matrix R_4 and sparse unitary matrices

We begin the algebraic analysis of the (4*4)-matrix R_4 in Fig. 5. Fig. 6 shows the decomposition of this matrix into a sum of 4 sparse matrices: $R_4 = R0_4 + R1_4 + R2_4 + R3_4$ (below we will explain that this decomposition is not arbitrary but constructed on the principle of dyadic-shift decompositions known in technology of digital signal processing).

$$
\begin{bmatrix} 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ -1 & -1 & 1 & 1 \end{bmatrix}
=
\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}
+
\begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{bmatrix}
+
\begin{bmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{bmatrix}
+
\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix}
$$

Fig. 6. The dyadic-shift decomposition of the matrix R_4 (Fig. 5) into the sum of 4 sparse matrices: $R_4 = R0_4 + R1_4 + R2_4 + R3_4$, where $R0_4$ is the identity matrix.

By definition, a complex square matrix U is unitary if its conjugate transpose $U^\dagger$ is also its inverse: $UU^\dagger = I$, where I is the identity matrix (the conjugate transpose $U^\dagger$ is also its inverse matrix U^{-1}). The real analogue of a unitary matrix is an orthogonal matrix, for which the conjugate transposition $U^\dagger$ is identical to the ordinary transposition: $UU^T = I$. In this article we consider only the case of real square matrix. Unitary matrices have significant importance in quantum mechanics because they preserve norms, and thus, probability amplitudes (https://en.wikipedia.org/wiki/Unitary_matrix). The tensor product of two unitary matrices always generates a unitary matrix [Rumer, Fet, 1970, p. 38].

It is interesting that each of sparse matrices $R0_4$, $R1_4$, $R2_4$ and $R3_4$ are unitary (or orthogonal since their entities are real):

$$R0_4 R0_4{}^T = I, \; R1_4 R1_4{}^T = I, \; R2_4 R2_4{}^T = I, \; R3_4 R3_4{}^T = I \qquad (1)$$

In molecular-genetic systems, relations of complementarity play the very important role. The book [Chapeville, Haenni, 1974, Chapter 1] notes that the proof of the complementary structure of the bases in DNA has led to the most fundamental discoveries in modern biology: this complementarity provides the most important properties of DNA as a carrier of genetic information, including DNA replication in the course of cell division and also all mechanisms of manifestation of genetic information. But one can note that the set of unitary genetic matrices $R0_4$, $R1_4$, $R2_4$ and $R3_4$ (Fig. 6) contains the following algebraic complementarities in their pairs: unitary matrices $R0_4$ and $R2_4$ form the first pair of the algebraic complementarity since they are transformed into each other by the mirror reflection relative to the average vertical line with simultaneous inversion of signs of their non-zero entries (the mirror-anti-symmetry). The same is true for unitary matrices $R1_4$ and $R3_4$, which form the second pair with the similar algebraic complementarity of the mirror-anti-symmetric type. The degeneracy of the genetic code is connected with such algebraic complementarities in the set of unitary genetic matrices.

Determinants of all the unitary matrices $R0_4$, $R1_4$, $R2_4$ and $R3_4$ are equal to 1; by this reason the matrices $R0_4$, $R1_4$, $R2_4$ and $R3_4$ belong to the type of so called special unitary matrices. The special unitary matrices are closed under multiplication and the inverse operation, and therefore form a matrix group called the special unitary group (http://mathworld.wolfram.com/SpecialUnitaryMatrix.html).

The table of multiplication of the closed set of genetic unitary matrices $R0_4$, $R1_4$, $R2_4$ and $R3_4$ is shown in Fig. 7. It coincides with the known multiplication table of the algebra of split-quaternions by J.Cockle, which are used in mathematics and physics [http://en.wikipedia.org/wiki/Split-quaternion ; Frenkel, Libine, 2011].

*	$R0_4$	$R1_4$	$R2_4$	$R3_4$
$R0_4$	$R0_4$	$R1_4$	$R2_4$	$R3_4$
$R1_4$	$R1_4$	$-R0_4$	$R3_4$	$-R2_4$
$R2_4$	$R2_4$	$-R3_4$	$R0_4$	$-R1_4$
$R3_4$	$R3_4$	$R2_4$	$R1_4$	$R0_4$

×	1	i	j	k
1	1	i	j	k
i	i	−1	k	−j
j	j	−k	1	−i
k	k	j	i	1

Fig. 7. On left: the multiplication table of the unitary genetic (4*4)-matrices $R0_4$, $R1_4$, $R2_4$ and $R3_4$ from Fig. 6. On right: the multiplication table of split-quaternions by J. Cockle (from

[http://en.wikipedia.org/wiki/Split-quaternion]).

From this fact, one can conclude that the division of the set of 16 doublets in line with the degeneracy of the genetic code is connected with the set of sparse unitary matrices R_{40}, R_{41}, R_{42} and R_{43}.

The rows of each of the unitary genetic matrices R_{04}, R_{14}, R_{24} and R_{34} form a complete orthogonal system of functions. The action of each of these matrices (except for the identity matrix R_{04}) on an arbitrary 4-dimensional vector X = [x_0, x_1, x_2, x_3] transforms it into a new vector Y, which can be considered as a spectral representation of the vector X on the basis of the orthogonal system of functions in the rows of the given matrix. The action of the same unitary matrix, taken in its transposed form, on this vector Y restores the original vector X. The exponentiation of each of the matrices R_{40}, R_{41}, R_{42} and R_{43} in a tensor power generates a new unitary matrix with an orthogonal system of functions in its rows and columns.

One can add that each of (2*2)-matrices in 4 quadrants of the genetic matrix R_4 is the sum of 2 unitary matrices. Really, the matrix in two quadrants along the main diagonal [1, -1; 1, 1] is the sum of two unitary matrices [1, 0; 0, 1] and [0, -1; 1, 0]; the matrix in two quadrants along the secondary diagonal is also the sum of two unitary matrices [1, 0; 0, -1] and [0, -1; -1, 0]. One of these unitary matrices is the well-known quantum gate Z=[1, 0; 0, -1].

Unitary matrices are used in quantum informatics as quantum logic elements (quantum gates) for performing quantum computations on their basis. In the case of multi-qubit systems, the operation of the tensor product of matrices is of key importance in connection with the postulate of quantum mechanics: the state space of a composite system is the tensor product of the state spaces of its components. In the light of this, it is especially interesting that the entire genetic (4x4)-matrix R_4 (Fig. 5) is constructed as the sum of the tensor products of four unitary (2*2)-matrices, that is, of four quantum gates U_0, U_1, U_2 and U_3 (Fig. 8) in line with the following expression (2):

$$R_4 = U_0 \otimes U_0 + U_0 \otimes U_1 + U_3 \otimes U_2 + U_3 \otimes (-U_3) \qquad (2)$$

where matrices U_0, U_1, U_2 and U_3 are shown in Fig. 8. These matrices are unitary: $U_0 * U_0^T = I_2$, $U_1 * U_1^T = I_2$, $U_2 * U_2^T = I_2$, $U_3 * U_3^T = I_2$, where I_2 is the identity matrix.

$$U_0 = \begin{vmatrix} 1, 0 \\ 0, 1 \end{vmatrix} ; \quad U_1 = \begin{vmatrix} 0, -1 \\ 1, 0 \end{vmatrix} ; \quad U_2 = \begin{vmatrix} -1, 0 \\ 0, 1 \end{vmatrix} ; \quad U_3 = \begin{vmatrix} 0, -1 \\ -1, 0 \end{vmatrix}$$

Fig. 8. Unitary (2*2)-matrices U_0, U_1, U_2 and U_3.

The set of these 4 matrices is also closed under multiplication. Fig. 9 shows their multiplication table, which coincides with the multiplication table of split-quaternions by J. Cockle by analogy with the case of unitary (4*4)-matrices in Figs. 6 and 7.

*	U_0	U_1	U_2	U_3
U_0	U_0	U_1	U_2	U_3
U_1	U_1	$-U_0$	U_3	$-U_2$
U_2	U_2	$-U_3$	U_0	$-U_1$
U_3	U_3	U_2	U_1	U_0

Fig. 9. The multiplication table of unitary (2*2)- matrices U_0, U_1, U_2 and U_3 from Fig. 8. It coincides with the multiplication table of split-quaternions by J. Cockle reproduced in Fig. 8.

It should be noted that unitary matrices U_0, U_1, U_2 and U_3 (Fig. 8) have relations with quantum gates used widely in quantum computing [Nielsen, Chuang, 2010, p. XXX].

Exponentiation of unitary matrices U_1, U_2 and U_3 into ordinary integer powers n = 2, 3, 4,… gives cyclic groups of matrices with the following periods: $U_1^n = U_1^{n+4}$, $U_2^n = U_2^{n+2}$, $U_3^n = U_3^{n+2}$. In this article we specially note a connection of cyclic groups with algebraic properties of genetic unitary matrices since such cyclic groups can be useful for modeling many inherited cyclic processes in physiology of organisms.

Exponentiation of each of unitary matrices U_0, U_1, U_2 and U_3 into tensor (or Kronecker) powers k = 2, 3, 4, … generates corresponding tensor families of unitary matrices: $U_0^{(k)}$, $U_1^{(k)}$, $U_2^{(k)}$ and $U_3^{(k)}$ where (k) means the tensor power.

4. The complementarity of sparse unitary matrices in genetics and the cruciform principle in inherited sensory informatics

This Section considers the cruciform character of the block black-and-white mosaic of the (4*4)-matrix R_4 of 16 doublets, which reflects essential peculiarities of the degeneracy of the genetic code (Figs. 4 and 5). One can note that genetically inherited constructions of physiological sensory-motor systems demonstrate similar cruciform structures. For example, the connection between the hemispheres of human brain and the halves of a human body possesses the similar cruciform character: the left hemisphere serves the right half of the body and the right hemisphere (Fig. 10) [Annett, 1985, 1992; Gazzaniga, 1995; Hellige, 1993]. The system of optic cranial nerves from two eyes possesses the cruciform structures as well: the optic nerves transfer information about the right half of field of vision into the left hemisphere of brain, and information about the left half of field of vision into the right hemisphere. The same is held true for the hearing system [Penrose, 1989, Chapter 9]. In particular, due to existence of such inter-complementary right and left parts in genetically inherited visual and hearing systems, a person has a stereoscopic perception of his environment. Now we show that a similar cruciform character, which is represented in the mosaic matrix of 16 doublets (Figs. 4, 5, 11), is connected with the following fact: this mosaic matrix is a sum of two sparse unitary matrices that are algebraic complementary to each other (they are mirror-anti-symmetric to each other by analogy with the left and right halves of a human body) and that can be considered as the right and left parts of the cruciform matrix R_4.
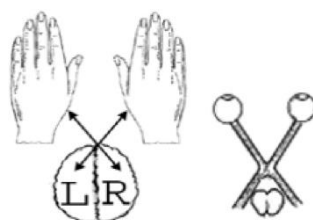
Fig. 10. The cruciform schemes of some morpho-functional structures of informational systems in human organism. On the left side: the cruciform connections of brain hemispheres with the left and the right halves of a human body. In the right side: the cruciform structure of optic nerves from eyes in brain.

One can suppose that this inherited cruciform character of sensory-motor systems is connected with genetic cruciform structures that include, in particular, the genetic matrices R_4 and R_8 [Petoukhov, 2008; Petoukhov, He, 2009]. Taking into account the quantum-informational character of molecular-genetic systems and also an important role of unitary matrices in quantum mechanics, it is interesting that – as we have discovered - these genetic matrices R_4 and R_8 are connected with inter-complementary sparse unitary matrices described below.

Let us begin with a consideration of the genetic matrix R_4 of 16 doublets. We reveal that its cruciform character is connected with a pair of two sparse unitary matrices, which are mirror-anti-symmetric to each other. Fig. 11 shows that the genetic cruciform matrix R_4 is the sum of two sparse matrices R_{4R} and R_{4L}: $R_4 = R_{4R}+R_{4L}$. These two sparse matrices are inter-complementary in an algebraic sense since they mirror-anti-symmetric to each other and they jointly form the non-sparse matrix R_4. Using the analogy with our stereoscopic vision by means of two – left and right – eyes, we conditionally

call the pair of complementary matrices R_{4R} and R_{4L} as the stereoscopic pair (or briefly, the stereo-pair) where the matrix R_{4R} is called the right stereo-matrix and the matrix R_{4L} is called the left stereo-matrix.

1	-1	1	-1
1	1	-1	-1
1	-1	1	-1
-1	-1	1	1

=

1	-1	0	0
1	1	0	0
0	0	1	-1
0	0	1	1

+

0	0	1	-1
0	0	-1	-1
1	-1	0	0
-1	-1	0	0

Fig. 11. The cruciform matrix R_4 is the sum of two sparse matrices R_{4R} and R_{4L}, non-zero entries of which coincide with non-zero entries in corresponding quadrants along the main diagonal and the secondary diagonal of the matrix R_4.

Taking with the factor $2^{-0.5}$, each of these stereo-matrices R_{4R} and R_{4L} is the unitary matrix:

$$(2^{-0.5}*R_{4R})*(2^{-0.5}*R_{4R})^T = I; \quad (2^{-0.5}*R_{4L})*(2^{-0.5}*R_{4L})^T = I. \tag{3}$$

Under actions of unitary matrices $2^{-0.5}*R_{4R}$ and $2^{-0.5}*R_{4L}$, an arbitrary 4-dimensional vector $\bar{X}$ = $[x_0, x_1, x_2, x_3]$ is transformed by the following manner:

$$(2^{-0.5}*R_{4R})*\bar{X}^T = 2^{-0.5}*[x_0-x_1, x_0+x_1, x_2-x_3, x_2+x_3]^T$$
$$(2^{-0.5}*R_{4L})*\bar{X}^T = 2^{-0.5}*[x_2-x_3, -x_2-x_3, x_0-x_1, -x_0-x_1]^T \tag{4}$$

In each of unitary matrices $2^{-0.5}*R_{4R}$ and $2^{-0.5}*R_{4L}$, the set of its rows contains a complete orthogonal system of functions. The multiplication of each of these sparse unitary matrices with an arbitrary 4-dimensional vector $\bar{Y}$ leads to a spectral representation $\bar{Z}$ of the vector $\bar{Y}$ on the basis of the complete orthogonal system of functions in rows of the unitary matrix. For example, in the case of the vector $\bar{Y}$ = [5, -3, 7, 9], its spectral representation $\bar{Z}_R$ on the basis of the system of functions in rows of the unitary matrix $2^{-0.5}*R_{4R}$ is the following ($\bar{Y}^T$ means the transposition of the vector $\bar{Y}$):

$$\bar{Z}_R = (2^{-0.5}*R_{4R})*\bar{Y}^T = 2^{-0.5}*[8, 2, -2, 16]^T \tag{5}$$

This spectral representation $\bar{Z}_R$ means that the vector $\bar{Y}$ is the sum of the following 4 basic vectors:
- the vector $2^{-0.5}*[1,-1,0,0]$ from the first row of the matrix $2^{-0.5}*R_{4R}$ (Fig. 11) multiplied by the factor $2^{-0.5}*8$;
- the vector $2^{-0.5}*[1,1,0,0]$ from the second row of $2^{-0.5}*R_{4R}$ multiplied by the factor $2^{-0.5}*2$;
- the vector $2^{-0.5}*[0,0,1,-1]$ from the third row of $2^{-0.5}*R_{4R}$ multiplied by the factor $-2^{-0.5}*2$;
- the vector $2^{-0.5}*[0,0,1,1]$ from the fourth row of $2^{-0.5}*R_{4R}$ multiplied by the factor $2^{-0.5}*16$.

Really, $2^{-0.5}*8*(2^{-0.5}*[1,-1,0,0]) + 2^{-0.5}*2*(2^{-0.5}*[1,1,0,0]) - 2^{-0.5}*2*(2^{-0.5}*[0,0,1,-1]) + 2^{-0.5}*16*(2^{-0.5}*[0,0,1,1]) = [5, -3, 7, 9] = \bar{Y}$.
The multiplication of the spectral vector $\bar{Z}_R^T$ with unitary transposed matrix $2^{-0.5}*R_{4R}^T$ restores automatically the initial vector $\bar{Y}$: $\bar{Z}_R^T*(2^{-0.5}*R_{4R}^T) = [5, -3, 7, 9]$.
Turning to the left stereo-matrix R_{4L}, one can check that the action of the unitary matrix $2^{-0.5}*R_{4L}$ on the same vector $\bar{Y}$ = [5, -3, 7, 9] leads to the following spectral representation $\bar{Z}_L$ of the vector $\bar{Y}$ on the basis of orthogonal functions in its rows:

$$\bar{Z}_L = (2^{-0.5}*R_{4R})*\bar{Y}^T = 2^{-0.5}*[-2, -16, \quad 8, \quad -2]^T \tag{6}$$

The multiplication of the spectral vector $\bar{Z}_L^T$ with unitary transposed matrix $2^{-0.5}*R_{4L}^T$ restores automatically the initial vector $\bar{Y}$: $\bar{Z}_L^T*(2^{-0.5}*R_{4L}^T)$ = [5, -3, 7, 9]. Such spectral representations (5, 6) of

vector-signals can be used for noise-immunity coding of information by some analogies with known noise-immunity coding on the basis of Hadamard matrices in digital signal processing.

Exponentiation of the unitary matrix $2^{-0.5}*R_{4L}$ into integer powers generates a cyclic group with the period 2: $(2^{-0.5}*R_{4L})^n = (2^{-0.5}*R_{4L})^{n+2}$ (the matrix R_{4L} specifies the transformation of reflection in the informational 4-dimensional space). In contrast, exponentiation of the unitary matrix $2^{-0.5}*R_{4R}$ in integer powers generates a cyclic group with the period 8: $(2^{-0.5}*R_{4R})^n = (2^{-0.5}*R_{4R})^{n+8}$ (the matrix $2^{-0.5}*R_{4R}$ specifies the transformation of turn in the informational 4-dimensional space). Fig. 12 shows transformations of the vector $\bar{X} = [x_0, x_1, x_2, x_3]$ under actions of the matrix $(2^{-0.5}*R_{4R})^n$.

$(2^{-0.5}*R_{4R})*\bar{X}^T =$ $2^{-0.5}*[x_0-x_1, x_0+x_1, x_2-x_3, x_2+x_3]^T$	$(2^{-0.5}*R_{4R})^5 *\bar{X}^T =$ $2^{-0.5}*[x_1-x_0, -x_0-x_1, x_3-x_2, -x_2-x_3]^T$
$(2^{-0.5}*R_{4R})^2 *\bar{X}^T = [-x_1, x_0, -x_3, x_2]^T$	$(2^{-0.5}*R_{4R})^6 *\bar{X}^T = [x_1, -x_0, x_3, -x_2]^T$
$(2^{-0.5}*R_{4R})^3 *\bar{X}^T =$ $2^{-0.5}*[-x_0-x_1, x_0-x_1, -x_2-x_3, x_2-x_3]^T$	$(2^{-0.5}*R_{4R})^7 *\bar{X}^T =$ $2^{-0.5}*[x_0+x_1, x_1-x_0, x_2+x_3, x_3-x_2]^T$
$(2^{-0.5}*R_{4R})^4 *\bar{X}^T = [-x_0, -x_1, -x_2, -x_3]^T$	$(2^{-0.5}*R_{4R})^8 *\bar{X}^T = [x_0, x_1, x_2, x_3]^T$
	$(2^{-0.5}*R_{4R})^9 *\bar{X}^T =$ $2^{-0.5}*[x_0-x_1, x_0+x_1, x_2-x_3, x_2+x_3]^T$

Fig. 12. Transformations of the vector $\bar{X} = [x_0, x_1, x_2, x_3]$ under actions of the matrix $(2^{-0.5}*R_{4R})^n$.

Another important difference between unitary matrices $2^{-0.5}*R_{4R}$ and $2^{-0.5}*R_{4L}$ is the following: the matrix $2^{-0.5}*R_{4L}$ is symmetric and correspondingly all its eigenvalues are real. By contrast, the matrix $2^{-0.5}*R_{4R}$ is asymmetric and has complex eigenvalues.

The multiplication of the stereo-matrices R_{4R} and R_{4L} are non-commutative: $R_{4R}*R_{4L} \neq R_{4L}*R_{4R}$. Their commutator $R_{4R}*R_{4L} - R_{4L}*R_{4R}$ taken with the factor $8^{-0.5}$ is the unitary matrix: $(R_{4R}*R_{4L}-R_{4L}*R_{4R})*(R_{4R}*R_{4L}-R_{4L}*R_{4R})^T/8 = I$.

In relation to each other, the right and left stereo-matrices R_{4R} and R_{4L} are amicable and disjoint (these algebraic notions are used in different applications of matrices). In linear algebra, by definition, two square matrices M and N of order n are said to be amicable if $MN^T = NM^T$ (see, for example, [Seberry, Wysocki, Wysocki, 2005]). Also by definition, two {0,±1} matrices M and N of the same size are said to be disjoint if for all of their positions the following rule is true: if M has a nonzero entry at the (i, j)-th position then N has zero entry at the same position and vice versa, i.e., M*N=0.

Stereo-matrices R_{4R} and R_{4L} taken in tensor power k = 1, 2, 3,… defines tensor families of matrices $R_{4R}^{(k)}$ and $R_{4L}^{(k)}$ of order 4^k, members of which satisfy the condition (6):

$$R_{4R}^{(k)}*(R_{4R}^{(k)})^T = I_s*4^{k/2} \ \text{ and } \ R_{4L}^{(k)}*(R_{4L}^{(k)})^T = I_s*4^{k/2} , \tag{7}$$

where I_s is the identity matrix of order 4^k.

The second tensor power of (4*4)-matrices R_{4R} and R_{4L}, that is (16*16)-matrices $R_{4R}^{(2)}$ and $R_{4L}^{(2)}$, can be written in line with the block algorithms in Fig. 13. (64*64)-matrices $R_{4R}^{(3)}$ and $R_{4L}^{(3)}$ can be similarly generated from $R_{4R}^{(2)}$ and $R_{4L}^{(2)}$ correspondingly.

$R_{4R}^{(2)} =$

R_{4R}	$- R_{4R}$	0	0
R_{4R}	R_{4R}	0	0
0	0	R_{4R}	$-R_{4R}$
0	0	R_{4R}	R_{4R}

; $R_{4L}^{(2)} =$

0	0	R_{4R}	$-R_{4R}$
0	0	$-R_{4R}$	$-R_{4R}$
R_{4R}	$-R_{4R}$	0	0
$-R_{4R}$	$-R_{4R}$	0	0

Fig. 13. The algorithmic construction of the second tensor power of genetic stereo-matrices R_{4R} and R_{4L} from Fig. 11.

Genetic stereo-matrices $R_{4R}^{(k)}$ and $R_{4L}^{(k)}$ have some analogies with Hadamard matrices. By definition, Hadamard matrices H_n of order n is a square matrix with entries +1 or -1 that satisfies the condition (8):

$$H_n H_n^T = nI_n \,, \qquad\qquad\qquad (8)$$

where I_n is the $n*n$ identity matrix. The mentioned analogies allow considering applications of stereo-matrices $R_{4R}{}^{(k)}$ and $R_{4L}{}^{(k)}$ in quantum informatics, noise-immune coding and recovering information in the presence of noise and interferences, and in some other fields in a parallel with traditional using Hadamard matrices there (a web search of bibliography of different applications of Hadamard matrices gives 44 thousands of publications in the period 1978-2005 years [Seberry, Wysocki, Wysocki, 2005]).

As we can judge, stereo-matrices R_{4R} and R_{4L} didn't meet previously in mathematical natural sciences. They were discovered in the analysis of molecular-genetic structures and they seem to be new interesting mathematical tools for genetic researches, quantum informatics and some other areas. The pair of complementary stereo-matrices R_{4R} and R_{4L} gives new materials to discussions existing from the ancient time about a role of binary principles "Yin-Yang", "left-right", "male-female", "odd-even" in organization of Nature (see, for example, a collection of facts in the book [Ivanov, 1978]).

Exponentiation of each of genetic unitary (4*4)-matrices $2^{-0.5}*R_{4R}$ and $2^{-0.5}*R_{4L}$ in tensor (Kronecker) powers k = 2, 3, 4, ... gives new tensor families of unitary matrices $(2^{-0.5}*R_{4R})^{(k)}$ and $(2^{-0.5}*R_{4L})^{(k)}$ of order 4^k with complete orthogonal systems of functions in their rows and columns. As known, exponentiation of matrices in tensor power can generate matrices with fractal structures [Gazale, 1999]. Fig. 14 shows examples of fractal structures inside tensor families of matrices $(2^{-0.5}*R_{4R})^{(k)}$ and $(2^{-0.5}*R_{4L})^{(k)}$.
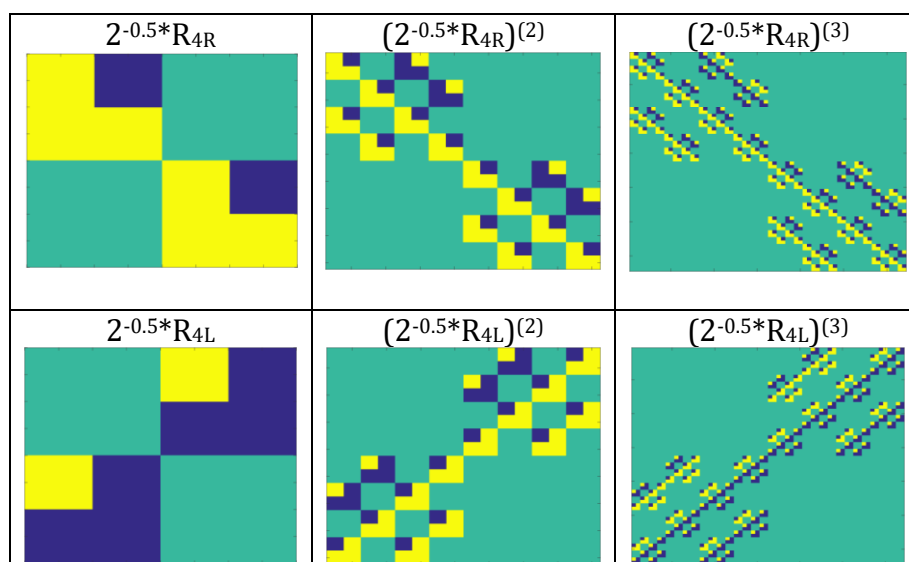
Fig. 14. Initial members of tensor families of matrices $(2^{-0.5}*R_{4R})^{(k)}$ and $(2^{-0.5}*R_{4L})^{(k)}$ having fractal structures. Entries +1, -1 and 0 are marked by yellow, blue and green correspondingly.

Let us note additionally that described decompositions of the matrix R_4 of 16 doublets (Figs. 4, 6, 11) into sparse unitary matrices were not arbitrary but they were based on objective binary-oppositional indicators of nitrogenous bases C, A, G, A and T/U (Fig. 1) and on the logical operation of modulo-2 addition. One can check that these decompositions were constructed in line with binary numbers in brackets inside matrix cells (Fig. 4). In each cell, such binary number is equal to a sum of binary numberings of the row and the column of the cell on the basis of modulo-2 addition. For example, the cell with the triplet CAG is located in the row 001 and in the column 011. The operation of modulo-2 addition gives their sum: 001⊕011 = 001; this binary number is shown in brackets in the cell with the triplet CAG. Such type of numeration of cells in matrices, whose rows and columns are numerated by means of dyadic groups of binary numbers, is known in theory of processing digital signals as a "dyadic-shift numeration" [Ahmed, Rao, 1975; Harmuth, 1989; Petoukhov, He, 2009]. Matrices with such numeration of their cells are called dyadic-shift matrices. One can see that such dyadic-shift numeration of cells of the (4*4)-matrix R_4 of 16 doublets (Fig. 4, left) divides the set of 16 doublets into 4 subsets with 4 doublets in each: the subset of cells with binary numberings 00 contains doublets CC, CG, GC and GG; the subset of cells with numberings 01 contains CA, CT, GA and GT; the subset of cells

with numberings 10 contains AC, AG, TC and TG; the subset of cells with numberings 11 contains AA, AT, TA and TT. Below we use similar dyadic-shift decompositions for the (8*8)-matrix R_8 of 64 triplets from Fig. 5 with non-trivial results. As known, if any system of elements demonstrates its connection with dyadic shifts, it indicates that the structural organization of its system is related to the logic of modulo-2 addition. Correspondingly the structural organization of the genetic system is related to the logic of modulo-2 addition.

One should note that each of two stereo-matrices R_{4R} and R_{4L} (Fig. 11) is the sum of two unitary matrices shown in Fig. 15: $R_{4R} = K_0+K_1$, $R_{4L} = K_2+K_3$. Really, $K_0*K_0^T =I_4$, $K_1*K_1^T =I_4$, $K_2*K_2^T =I_4$ and $K_3*K_3^T =I_4$, where I_4 is the identity matrix of order 4 and the matrix K_0 resembles the well-known quantum gate "controlled-NOT": [1, 0, 0, 0; 0, 1, 0, 0; 0, 0, 0, 1; 0, 0, 1, 0] [Nielsen, Chuang, 2010, p. XXXI].

Rows and columns of sparse matrices K_0, K_1, K_2 and K_3 correspond to complete orthogonal systems of functions. Below we will also meet other examples that genetic unitary matrices with block structures are constructed as sums of more simple unitary matrices.

$$R_{4R} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = K_0 + K_1$$

$$R_{4L} = \begin{bmatrix} 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & -1 \\ 1 & -1 & 0 & 0 \\ -1 & -1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix} = K_2 + K_3$$

Fig. 15. Each of stereo-matrices R_{4R} and R_{4L} (Fig. 11) is the sum of two unitary matrices: $R_{4R} = K_0+K_1$, $R_{4L} = K_2+K_3$, where K_0, K_1, K_2 and K_3 are unitary matrices.

Exponentiations of genetic unitary (4*4)-matrices K_0, K_1, K_2 and K_3 (Fig. 15) in tensor (Kronecker) powers n = 2, 3, 4, ... give tensor families of unitary matrices $K_0^{(n)}$, $K_1^{(n)}$, $K_2^{(n)}$ and $K_3^{(n)}$ of order 4^n with complete orthogonal systems of functions in their rows and columns. Fig. 16 shows initial members of these tensor families with their fractal structures, which are constituent parts of the fractal structures in Fig. 14.
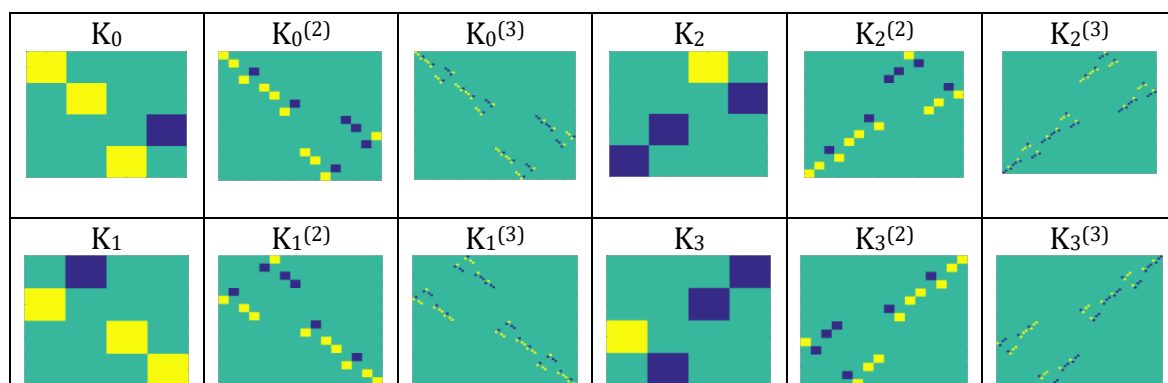
Fig. 16. Initial members of tensor families of unitary matrices $K_0^{(n)}$, $K_1^{(n)}$, $K_2^{(n)}$ and $K_3^{(n)}$ having fractal structures and complete orthogonal systems of functions in their rows and columns. Entries +1, -1 and 0 are marked by yellow, blue and green correspondingly.

5. The 8 sparse unitary matrices in the dyadic-shift decomposition of genetic matrix R_8 of 64 triplets

Turn now to the algebraic analysis of the (8*8)-matrix R_8 (Fig. 5). Fig. 17 shows the dyadic-shift decomposition of the matrix R_8 into 8 sparse unitary matrices: $R_8 = R_{80} + R_{81} + R_{82} + R_{83} + R_{84} + R_{85} + R_{86} + R_{87}$.

One can note that the set of 8 unitary genetic matrices R_{80}, R_{81}, R_{82}, R_{83}, R_{84}, R_{85}, R_{86}, R_{87} (Fig. 17) also contains the following algebraic complementarities in corresponding pairs of these matrices: unitary matrices R_{80} and R_{87} form the first pair of the algebraic complementarity since they are transformed into each other by mirror reflection relative to the average vertical line with inversion of signs of their non-zero entries (the mirror-anti-symmetry). The same is true for the pairs of unitary matrices R_{81} and R_{86}, R_{82} and R_{85}, R_{83} and R_{84}, which form the other pairs with their similar algebraic complementarity of the mirror-anti-symmetric type.

$$
R_8 =
\begin{bmatrix}
1&0&0&0&0&0&0&0\\
0&1&0&0&0&0&0&0\\
0&0&1&0&0&0&0&0\\
0&0&0&1&0&0&0&0\\
0&0&0&0&1&0&0&0\\
0&0&0&0&0&1&0&0\\
0&0&0&0&0&0&1&0\\
0&0&0&0&0&0&0&1
\end{bmatrix}
+
\begin{bmatrix}
0&1&0&0&0&0&0&0\\
1&0&0&0&0&0&0&0\\
0&0&0&1&0&0&0&0\\
0&0&1&0&0&0&0&0\\
0&0&0&0&0&1&0&0\\
0&0&0&0&1&0&0&0\\
0&0&0&0&0&0&0&1\\
0&0&0&0&0&0&1&0
\end{bmatrix}
+
\begin{bmatrix}
0&0&-1&0&0&0&0&0\\
0&0&0&-1&0&0&0&0\\
1&0&0&0&0&0&0&0\\
0&1&0&0&0&0&0&0\\
0&0&0&0&0&0&-1&0\\
0&0&0&0&0&0&0&-1\\
0&0&0&0&1&0&0&0\\
0&0&0&0&0&1&0&0
\end{bmatrix}
+
\begin{bmatrix}
0&0&0&-1&0&0&0&0\\
0&0&-1&0&0&0&0&0\\
0&1&0&0&0&0&0&0\\
1&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&-1\\
0&0&0&0&0&0&-1&0\\
0&0&0&0&0&1&0&0\\
0&0&0&0&1&0&0&0
\end{bmatrix}
$$

$$
+
\begin{bmatrix}
0&0&0&0&1&0&0&0\\
0&0&0&0&0&1&0&0\\
0&0&0&0&0&0&-1&0\\
0&0&0&0&0&0&0&-1\\
1&0&0&0&0&0&0&0\\
0&1&0&0&0&0&0&0\\
0&0&-1&0&0&0&0&0\\
0&0&0&-1&0&0&0&0
\end{bmatrix}
+
\begin{bmatrix}
0&0&0&0&0&1&0&0\\
0&0&0&0&1&0&0&0\\
0&0&0&0&0&0&0&-1\\
0&0&0&0&0&0&-1&0\\
0&1&0&0&0&0&0&0\\
1&0&0&0&0&0&0&0\\
0&0&0&-1&0&0&0&0\\
0&0&-1&0&0&0&0&0
\end{bmatrix}
+
\begin{bmatrix}
0&0&0&0&0&0&-1&0\\
0&0&0&0&0&0&0&-1\\
0&0&0&-1&0&0&0&0\\
0&0&0&0&-1&0&0&0\\
0&0&-1&0&0&0&0&0\\
0&0&0&-1&0&0&0&0\\
-1&0&0&0&0&0&0&0\\
0&-1&0&0&0&0&0&0
\end{bmatrix}
+
\begin{bmatrix}
0&0&0&0&0&0&0&-1\\
0&0&0&0&0&0&-1&0\\
0&0&0&0&0&-1&0&0\\
0&0&0&0&-1&0&0&0\\
0&0&0&-1&0&0&0&0\\
0&0&-1&0&0&0&0&0\\
0&-1&0&0&0&0&0&0\\
-1&0&0&0&0&0&0&0
\end{bmatrix}
$$

Fig. 17. The dyadic-shift decomposition of the matrix R_8 into 8 sparse unitary matrices: $R_8 = R_{80} + R_{81} + R_{82} + R_{83} + R_{84} + R_{85} + R_{86} + R_{87}$.

Each of these 8 sparse matrices is unitary: $R_{80}*R_{80}^T = I$, $R_{81}*R_{81}^T = I$, $R_{82}*R_{82}^T = I$, $R_{83}*R_{83}^T = I$, $R_{84}*R_{84}^T = I$, $R_{85}*R_{85}^T = I$, $R_{86}*R_{86}^T = I$, $R_{87}*R_{87}^T = I$.

Determinants of all these 8 unitary matrices are equal to 1; by this reason they belong to the type of so called special unitary matrices. The special unitary matrices are closed under multiplication and the inverse operation. They form the special unitary group (https://en.wikipedia.org/wiki/Special_unitary_group). The multiplication table (Fig. 18) of this closed set of 8 unitary matrices coincides with the multiplication table of the algebra of bi-split-quaternions of Cockle.

The set of these 8 sparse matrices is divided into 2 equal subsets with 4 matrices in each. The first subset contains matrices R_{80}, R_{81}, R_{82}, R_{83}, whose non-zero entries belong only to two quadrants along the main diagonal of the matrix R_8 (in the symbolic matrix of triplets in Fig. 4, all triplets in these two quadrants begin with letters C or G). In this subset two matrices R_{80} and R_{81} are symmetric and two other matrices R_{82} and R_{83} are asymmetric. The second subset contains matrices R_{84}, R_{85}, R_{86}, R_{87}, whose non-zero entries belong only to two quadrants along the second diagonal of the matrix R_8 (in the symbolic matrix of triplets in Fig. 4, all triplets in these two quadrants begin with letters A or T). All these 4 matrices are symmetric and correspondingly have real eigenvalues.

*	R_{80}	R_{81}	R_{82}	R_{83}	R_{84}	R_{85}	R_{86}	R_{87}

R_{80}	R_{80}	R_{81}	R_{82}	R_{83}	R_{84}	R_{85}	R_{86}	R_{87}
R_{81}	R_{81}	R_{80}	R_{83}	R_{82}	R_{85}	R_{84}	R_{87}	R_{86}
R_{82}	R_{82}	R_{83}	$-R_{80}$	$-R_{81}$	$-R_{86}$	$-R_{87}$	R_{84}	R_{85}
R_{83}	R_{83}	R_{82}	$-R_{81}$	$-R_{80}$	$-R_{87}$	$-R_{86}$	R_{85}	R_{84}
R_{84}	R_{84}	R_{85}	R_{86}	R_{87}	R_{80}	R_{81}	R_{82}	R_{83}
R_{85}	R_{85}	R_{84}	R_{87}	R_{86}	R_{81}	R_{80}	R_{83}	R_{82}
R_{86}	R_{86}	R_{87}	$-R_{84}$	$-R_{85}$	$-R_{82}$	$-R_{83}$	R_{80}	R_{81}
R_{87}	R_{87}	R_{86}	$-R_{85}$	$-R_{84}$	$-R_{83}$	$-R_{82}$	R_{81}	R_{80}

Fig. 18. The multiplication table of 8 genetic unitary matrices R_{80}, R_{81}, R_{82}, R_{83}, R_{84}, R_{85}, R_{86}, R_{87} from Fig. 17 coincides with the multiplication table of the algebra of bi-split-quaternions of Cockle.

The exponentiation of the sum of matrices R_{80}, R_{81}, R_{82} and R_{83} (Fig. 17) taken with the factor $8^{-0.5}$ defines a cyclic group with its period 8: $(8^{-0.5}*(R_{80}+R_{81}+R_{82}+R_{83}))^n = (8^{-0.5}*(R_{80}+R_{81}+R_{82}+R_{83}))^{n+8}$.

The exponentiation of the sum of matrices R_{84}, R_{85}, R_{86} and R_{87} (Fig. 17) taken with the factor $8^{-0.5}$ defines a cyclic group with its period 2: $(8^{-0.5}*(R_{84}+R_{85}+R_{86}+R_{87}))^n = (8^{-0.5}*(R_{84}+R_{85}+R_{86}+R_{87}))^{n+8}$. Correspondingly this matrix $8^{-0.5}*(R_{84}+R_{85}+R_{86}+R_{87})$ defines transformation of reflection in the 8-dimensional informational space.

One should note that these two subsets of 8 unitary genetic matrices have some relations to evolution changes of the genetic code. Modern science knows more than 20 variants (or dialects) of the genetic code represented on website http://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi. As this website shows, these dialects differ from the Standard Code only by the code meanings of a small number of triplets. For example, in the Standard Code the triplet TAG encodes the amino acid Leu but in the Chlorophycean Mitochondrial Code it serves as the stop-codon. One can check that in all dialects without exception all stop-codons have their first letter A or T; practically all initiative codons also begin with the letters A or T. Beside this, practically all triplets with changed coding meaning are triplets with the first letters A or T in them; such triplets belong to two quadrants along the second diagonal of the matrix in Fig. 4. All entries of these two quadrants in the corresponding numeric matrix in Fig. 5 belong to the second subset of symmetric unitary matrices R_{84}, R_{85}, R_{86} and R_{87} (Fig. 17). A small exception is the case of two genetic codes of yeast (the Yeast Mitochondrial Code and the Alternative Yeast Nuclear Code), in which the triplets CUU, CUC, CUA and CUG, having the first letter C, change their coding meaning: in the Standard Code, all these 4 triplets encode the amino acid Leu but in the Yeast Mitochondrial Code they encode the amino acid Thr. These 4 triplets having strong roots are located in the upper quadrant of the matrix of triplets in Fig. 4, and their representations by entries +1 in the numeric matrix R_8 belong to the unitary matrices R_{82} and R_{83} (Fig. 17).

The genetic (8*8)-matrix R_8 can be represented via unitary genetic (2*2)-matrices U_0, U_1, U_2 and U_3 from Fig. 8:

$$R_8 = U_0 \otimes U_0 \otimes U_0 + U_0 \otimes U_0 \otimes (-U_3) + U_0 \otimes U_1 \otimes U_0 + U_0 \otimes U_1 \otimes (-U_3) +$$
$$U_3 \otimes U_2 \otimes U_0 + U_3 \otimes U_2 \otimes (-U_3) + U_3 \otimes (-U_3) \otimes U_0 + U_3 \otimes (-U_3) \otimes (-U_3) \qquad (9)$$

Each of the 8 summands on the right-hand side of the expression (9) coincides with one of the above unitary matrices R_{80}, R_{81}, R_{82}, R_{83}, R_{84}, R_{85}, R_{86}, R_{87} (Fig. 17).

5. Connections among amino acids and triplets from the standpoint of unitary genetic matrices

Our results about connections of the genetic (8*8)-matrix R_8 (Fig. 5) with unitary matrices give additional approaches to study symmetric relations between the set of 64 triplets and the set of 20 amino acids and stop-codons encoded by triplets. Fig. 19 reproduces the symbolic matrix of 64 triplets from Fig. 4 but with the additional indication of amino acids and stop-codons in the Vertebrate Mitochondrial Code; this dialect is the most symmetrical among all dialects of the genetic code in line with known data of the website http://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi. One can see in Figs. 4, 5 and 18 that the genetic symbolic matrix of 64 triplets and its corresponding numeric representation R_8 (Fig. 5) have pairs of adjacent rows 0 and 1, 2 and 3, 4 and 5, 6 and 7, whose mosaics are identical inside each of pairs. Moreover, both rows inside each of these pairs have the identical list of amino acids and stop-codons marked by identical colors (Fig. 19).

	000	001	010	011	100	101	110	111
000	CCC Pro	CCA Pro	CAC His	CAA Gln	ACC Thr	ACA Thr	AAC Asn	AAA Lys
001	CCT Pro	CCG Pro	CAT His	CAG Gln	ACT Thr	ACG Thr	AAT Asn	AAG Lys
010	CTC Leu	CTA Leu	CGC Arg	CGA Arg	ATC Ile	ATA Met	AGC Ser	AGA Stop
011	CTT Leu	CTG Leu	CGT Arg	CGG Arg	ATT Ile	ATG Met	AGT Ser	AGG Stop
100	TCC Ser	TCA Ser	TAC Tyr	TAA Stop	GCC Ala	GCA Ala	GAC Asp	GAA Glu
101	TCT Ser	TCG Ser	TAT Tyr	TAG Stop	GCT Ala	GCG Ala	GAT Asp	GAG Glu
110	TTC Phe	TTA Leu	TGC Cys	TGA Trp	GTC Val	GTA Val	GGC Gly	GGA Gly
111	TTT Phe	TTG Leu	TGT Cys	TGG Trp	GTT Val	GTG Val	GGT Gly	GGG Gly

Fig. 19. The correspondence among 64 triplets and 20 amino acids and stop-codons in the case of the Vertebrate Mitochondrial Code. Both rows inside each of pairs of adjacent rows 0-1, 2-3, 4-5, 6-7 with their even-odd numberings have identical black-and-white mosaics (reproduced in the numeric matrix R_8 in Fig. 5) and also identical lists of amino acids and stop-codons (marked with the same colors).

The described decomposition (Fig. 17) of the numeric matrix R_8 of triplets into sums of unitary matrices is accompanied by a separation of the complete set of 64 triplets into corresponding subsets of triplets. This separation of the set of 64 triplets is accompanied by a relevant separation of the complete set of 20 amino acids and stop–codons into appropriate subsets taking into account the code meaning of each of triplets. For example, the matrix R_8 can be represented in the following form using the decomposition in Fig. 17:

$$R_8 = (R_{80} + R_{82}) + (R_{81} + R_{83}) + (R_{84} + R_{86}) + (R_{85} + R_{87}) \tag{10}$$

Here each of 4 expressions in brackets taken with the factor $2^{-0.5}$ is a unitary matrix:

$$0.5*(R_{80}+R_{82})(R_{80}+R_{82})^T = I_8; \quad 0.5*(R_{81}+R_{83})(R_{81}+R_{83})^T = I_8;$$
$$0.5*(R_{84}+R_{86})(R_{84}+R_{86})^T = I_8; \quad 0.5*(R_{85}+R_{87})(R_{85}+R_{87})^T = I_8 \tag{11}$$

where I_8 is the identity matrix of order 8.

Fig. 20 shows these 4 inter-complementary matrices $(R_{80}+R_{82})$, $(R_{81}+R_{83})$, $(R_{84}+R_{86})$ and $(R_{85}+R_{87})$ whose sum is equal to the matrix R_8 from Fig. 5.

$$\begin{array}{cccc|cccc} 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \end{array} \qquad \begin{array}{cccc|cccc} 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \end{array}$$

$$R_{80}+R_{82} = \begin{array}{cccc|cccc} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{array} \quad ; \quad R_{81}+R_{83} = \begin{array}{cccc|cccc} 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{array}$$

$$R_{84}+R_{86} = \begin{array}{cccc|cccc} 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & -1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & -1 & 0 & 0 & 0 & 0 \end{array} \quad ; \quad R_{85}+R_{87} = \begin{array}{cccc|cccc} 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 \\ 0 & 0 & 0 & 0 & -1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & -1 & 0 & 0 & 0 & 0 \\ -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \end{array}$$

Fig. 20. The result of the decomposition of the genetic matrix R_8 (Fig. 5) into 4 inter-complementary matrices $(R_{80}+R_{82})$, $(R_{81}+R_{83})$, $(R_{84}+R_{86})$ and $(R_{85}+R_{87})$, which, taken with the factor $2^{-0.5}$, are unitary matrices with their own complete systems of orthogonal functions. Matrices R_{80}, R_{81}, R_{82}, R_{83}, R_{84}, R_{85}, R_{86}, R_{87} are shown in Fig. 17.

Rows and columns of each of the 4 unitary matrices $2^{-0.5}*(R_{80}+R_{82})$, $2^{-0.5}*(R_{81}+R_{83})$, $2^{-0.5}*(R_{84}+R_{86})$ and $2^{-0.5}*(R_{85}+R_{87})$ represent complete systems of orthogonal functions. By this reason, these 4 matrices can be used for a decomposition of an arbitrary 8-dimensional vector $\bar{V}=[x_0, x_1, x_2, x_3, x_4, x_5, x_6, x_7]$ on the basis of orthogonal functions of each of these orthogonal systems. By this way, an 8-dimensional vector $\bar{V}$ receives different spectral representations in a general case. The multiplication of any of such spectral representations of the vector $\bar{V}$ with correspondence transposed unitary matrix restores the initial vector $\bar{V}$. Such spectral representations of vector-signals can be used for noise-immunity coding of information by some analogies with known noise-immunity coding on the basis of Hadamard matrices in digital signal processing.

Determinants of all the unitary matrices $2^{-0.5}*(R_{80}+R_{82})$, $2^{-0.5}*(R_{81}+R_{83})$, $2^{-0.5}*(R_{84}+R_{86})$ and $2^{-0.5}*(R_{85}+R_{87})$ are equal to 1. Some multiplications of the matrices P_{R0}, P_{R1}, P_{L0} and P_{L1} each with other are commutative and others are non-commutative (12):

$$(R_{80}+R_{82})(R_{81}+R_{83}) = (R_{81}+R_{83})(R_{80}+R_{82});$$
$$(R_{84}+R_{86})(R_{85}+R_{87}) = (R_{85}+R_{87})(R_{84}+R_{86});$$
$$(R_{80}+R_{82})(R_{84}+R_{86}) \neq (R_{84}+R_{86})(R_{80}+R_{82}) ;$$
$$(R_{80}+R_{82})(R_{85}+R_{87}) \neq (R_{85}+R_{87})(R_{80}+R_{82});$$
$$(R_{81}+R_{83})(R_{84}+R_{86}) \neq (R_{84}+R_{86})(R_{81}+R_{83});$$
$$(R_{81}+R_{83})(R_{85}+R_{87}) \neq (R_{85}+R_{87})(R_{81}+R_{83}) \tag{12}$$

Exponentiation of unitary matrices $2^{-0.5}*(R_{80}+R_{82})$ and $2^{-0.5}*(R_{81}+R_{83})$ defines cyclic groups with their period 8: $(2^{-0.5}*(R_{80}+R_{82}))^n = (2^{-0.5}*(R_{80}+R_{82}))^{n+8}$; $(2^{-0.5}*(R_{81}+R_{83}))^n = (2^{-0.5}*(R_{81}+R_{83}))^{n+8}$. Exponentiation of unitary matrices $2^{-0.5}*(R_{84}+R_{86})$ and $2^{-0.5}*(R_{85}+R_{87})$ defines cyclic groups with their period 2: $(2^{-0.5}*(R_{84}+R_{86}))^n = (2^{-0.5}*(R_{84}+R_{86}))^{n+2}$; $(2^{-0.5}*(R_{85}+R_{87}))^n = (2^{-0.5}*(R_{85}+R_{87}))^{n+2}$.

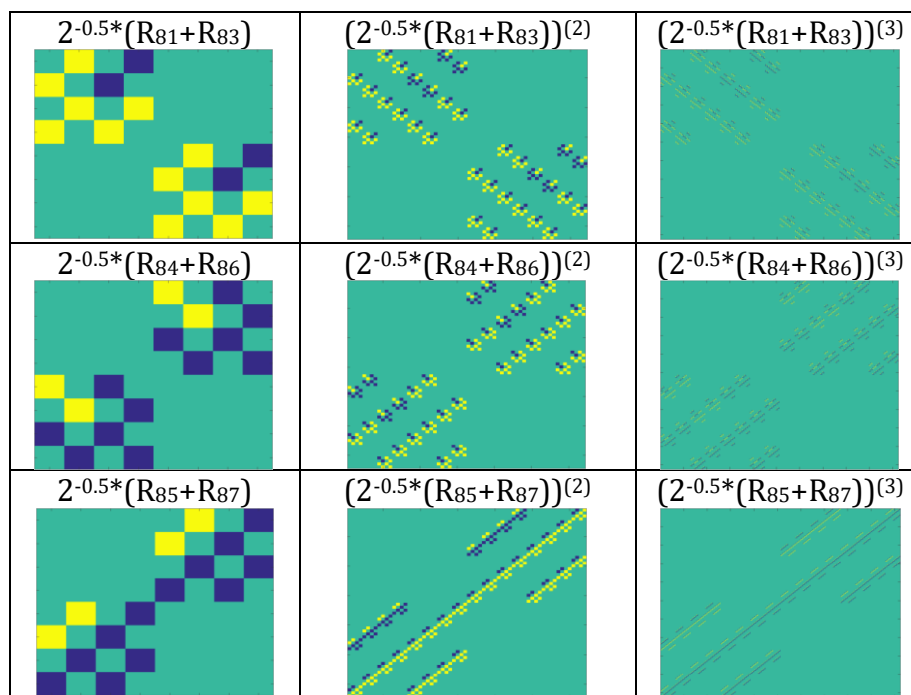| $2^{-0.5}*(R_{80}+R_{82})$	$(2^{-0.5}*(R_{80}+R_{82}))^{(2)}$	$(2^{-0.5}*(R_{80}+R_{82}))^{(2)}$
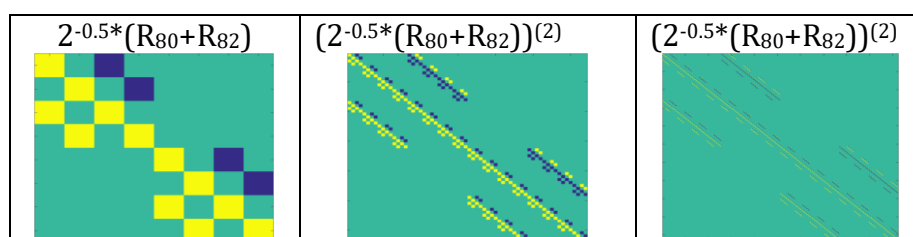
Fig. 21. Initial members of tensor families of unitary matrices $(2^{-0.5}*(R_{80}+R_{82}))^{(k)}$, $(2^{-0.5}*(R_{81}+R_{83}))^{(k)}$, $(2^{-0.5}*(R_{84}+R_{86}))^{(k)}$ and $(2^{-0.5}*(R_{85}+R_{87}))^{(k)}$ having fractal structures and complete orthogonal systems of functions. Entries $+2^{-0.5}$, $-2^{-0.5}$ and 0 are marked by yellow, blue and green correspondingly.

Exponentiation of genetic unitary (8*8)-matrices $2^{-0.5}*(R_{80}+R_{82})$, $2^{-0.5}*(R_{81}+R_{83})$, $2^{-0.5}*(R_{84}+R_{86})$ and $2^{-0.5}*(R_{85}+R_{87})$ in tensor (Kronecker) powers k = 2, 3, 4, … gives tensor families of new unitary matrices of order 8^k with new complete systems of orthogonal functions: $(2^{-0.5}*(R_{80}+R_{82}))^{(k)}$, $(2^{-0.5}*(R_{81}+R_{83}))^{(k)}$, $(2^{-0.5}*(R_{84}+R_{86}))^{(k)}$ and $(2^{-0.5}*(R_{85}+R_{87}))^{(k)}$ (Fig. 21).

The 4 matrices $(R_{80}+R_{82})$, $(R_{81}+R_{83})$, $(R_{84}+R_{86})$ and $(R_{85}+R_{87})$ are pairwise amicable: $(R_{80}+R_{82})(R_{81}+R_{83})^T = (R_{81}+R_{83})(R_{80}+R_{82})^T$; $(R_{80}+R_{82})(R_{84}+R_{86})^T = (R_{84}+R_{86})(R_{80}+R_{82})^T$; $(R_{80}+R_{82})(R_{85}+R_{87})^T = (R_{85}+R_{87})(R_{80}+R_{82})^T$; $(R_{81}+R_{83})(R_{84}+R_{86})^T = (R_{84}+R_{86})(R_{81}+R_{83})^T$; $(R_{81}+R_{83})(R_{85}+R_{87})^T = (R_{85}+R_{87})(R_{81}+R_{83})^T$.

Matrices $(R_{80}+R_{82})$, $(R_{81}+R_{83})$, $(R_{84}+R_{86})$ and $(R_{85}+R_{87})$ taken in tensor powers k = 1, 2, 3,… define corresponding tensor families whose matrices satisfy the condition (13):

$$(R_{80}+R_{82})^{(k)}*((R_{80}+R_{82})^{(k)})^T =I_w*4^{k/2}, \quad (R_{81}+R_{83})^{(k)}*((R_{81}+R_{83})^{(k)})^T =I_w*4^{k/2},$$
$$(R_{84}+R_{86})^{(k)}*((R_{84}+R_{86})^{(k)})^T= I_w*4^{k/2}, \quad (R_{85}+R_{87})^{(k)}*((R_{85}+R_{87})^{(k)})^T =I_w*4^{k/2} \qquad (13)$$

where I_w is the identity matrix of order 8^k.

Now let us comparise the distribution of amino acids and stop-codons in the symbolic matrix in Fig. 19 with the distribution of non-zero entries in numeric matrices $R_{80}+R_{82}$, $R_{81}+R_{83}$, $R_{84}+R_{86}$ and $R_{85}+R_{87}$ (Fig. 20). The comparison shows that the phenomenon of pairwise distributions of amino acids and stop-codons in adjacent rows 0-1, 2-3, 4-5 and 6-7 of the genetic matrix in Fig. 19 is connected with the set of 4 complementary unitary matrices $R_{80}+R_{82}$, $R_{81}+R_{83}$, $R_{84}+R_{86}$ and $R_{85}+R_{87}$. More precisely, these 4 matrices define the symmetric separations of the set of 64 triplets and the set of 20 amino acids and stop-codons into the following subsets:

- the set of non-zero entries in the matrix $R_{80}+R_{82}$ corresponds to the set of 10 amino acids Pro, His, Gln, Leu, Arg, Ala, Asp, Glu, Val, Gly, which are encoded by the set of 16 triplets - CCC, CAC, CCG, CAG, CTC, CGC, CTG, CGG, GCC, GAC, GCG, GAG, GTC, GGC, GTG and GGG;
- the set of non-zero entries in the matrix $R_{81}+R_{83}$ corresponds to the same set of 10 amino acids Pro, His, Gln, Leu, Arg, Ala, Asp, Glu, Val, Gly, which are encoded by another set of 16 triplets - CCA, CAA, CCT, CAT, CTA, CGA, CTT, CGT, GCA, GAA, GCT, GAT, GTA, GGA, GTT, GGT;
- the set of non-zero entries in the matrix $R_{84}+R_{86}$ corresponds to the following set of amino acids and stop-codons: Thr, Asn, Lys, Ile, Met, Ser, Tyr, Phe, Leu, Cys, Trp and 2 stop-codons, which are

encoded by the set of 16 triplets – ACC, AAC, ACG, AAG, ATC, AGC, ATG, AGG, TCC, TAC, TCG, TAG, TTC, TGC, TTG and TGG;

- the set of non-zero entries in the matrix $R_{85}+R_{87}$ corresponds to the same set of amino acids and stop-codons: Thr, Asn, Lys, Ile, Met, Ser, Tyr, Phe, Leu, Cys, Trp and 2 stop-codons, which are encoded by another set of 16 triplets – ACA, AAA, ACG, AAG, ATC, AGC, ATG, AGG, TCA, TAA, TCT, TAT, TTA, TGA, TTT and TGT.

The 32 triplets in both matrices $R_{80}+R_{82}$ and $R_{81}+R_{83}$ begin with letters C or G and define the content of two quadrants along the main diagonal of the symbolic matrix in Figs. 2, 19. Other 32 triplets in both matrices $R_{84}+R_{86}$ and $R_{85}+R_{87}$ begin with letters A and T and define the content of two quadrants along the secondary diagonal of the same matrix. It was mentioned above that in all dialects of the genetic code without exception all stop-codons have their first letters A or T; practically all initiative codons also begin with the letters A or T (so they are connected with unitary matrices $2^{-0.5}*(R_{84}+R_{86})$ and $2^{-0.5}*(R_{85}+R_{87})$). Beside this, practically all triplets with changed coding meaning in different dialects are also triplets with the first letters A or T in them and they are connected with the same unitary matrices.

One can note that our study belongs to the field of applications of algebraic methods and notions of quantum mechanics for modeling molecular-genetic structures where works of many authors exist, for example, the following [Dragovich, Dragovich, 2007; Dragovich, Khrennikov, Misic, 2017; Fimmel, Danielli, Strüngmann, 2013; Fimmel, Strüngmann, 2016; Igamberdiev, 1993; Ji, 2015, 2017; Hu, Petoukhov, 2017; Hu, Petoukhov, Petukhova, 2017; Matsuno, 1999, 2003, 2015; Matsuno, Paton, 2000; Moon Ho Lee et al, 2017; Pellionis, 2008; Pellionisz et al, 2012; Penrose, 1996; Perez, 2010, 2013, 2017; Petoukhov, 2008, 2010a,b, 2011, 2012, 2015a,b,c,d,c, 2016, 2017a,b; Petoukhov, He, 2009; Petoukhov, Petukhova, 2017a,b; Petoukhov, Svirin, 2012; Petoukhov et al, 2017; Rapoport, 2016a,b,c; Rumer, 1968; Simeonov, 2013; Stambuk, 1999].

6. On unitary genetic matrices and the law of frequencies of hydrogen bonds in long DNA sequences of complete sets of nuclear chromosomes

In this Section we will study hidden regularities in long sequences of single stranded DNA, which usually studied by some authors in relation with the second Chargaff's parity rule (see for example [Chargaff, 1971, 1975; Prahbu, 1993; Rapoport, Trifonov, 2012; https://en.wikipedia.org/wiki/Chargaff%27s_rules]. The term "long DNA sequences" means that each of sequences contains no less than 100000 nucleotides. We will consider any of long DNA sequences as a chain of numbers 2 and 3 of hydrogen bonds, which connect complementary nucleotides A-T and C-G correspondingly (one can denote conditionally A=T=2 and C=G=3). We call such chains of numbers of hydrogen bonds as "hydrogen bonds sequences" or briefly "H-sequences".

Each of long chains (or sequences) of numbers 2 and 3 of hydrogen bonds (for example, the H-sequence 3-2-3-3-3-2-2-2-3-2-3-3-…) can be represented also as a sequence of H-doublets (32-33-32-22-32-33-…), or a sequence of H-triplets (323-332-223-233-…) or, in a general case, as a sequence of H-n-plets (n = 1, 2, 3, 4, 5,…). We call such different representations of a long DNA sequence of nucleotides A, T, C, G in forms of sequences of H-n-plets as its "H-n-plet representations".

A complete set of H-n-plets under fixed value n is called "the alphabet of H-n-plets" and it contains 2^n members: for example, the alphabet of H-monoplets contains 2 members (3 and 2); the alphabet of H-doublets contains 4 members (33, 32, 23, 22); the alphabet of H-triplets contains $2^3 = 8$ members (333, 332, 323, 322, 233, 232, 223, 222); etc. One can mention that the tensor family of vectors $[3, 2]^{(n)}$ (where (n) means the tensor power, n = 1, 2, 3, 4, …) contains vectors, whose sets of components coincide with corresponding alphabets of H-n-plets. For example, $[3, 2]^{(2)} = [33, 32, 23, 22]$; $[3, 2]^{(3)} = [333, 332, 323, 322, 233, 232, 223, 222]$, etc. By this reason, one can note conditionally that all H-n-plets alphabets belong to the same tensor family of vectors and are interconnected each other.

Below we describe our discovery of those hidden regularities in H-sequences of long DNA and in H-sequences of complete sets of chromosomes of different organisms, which are connected with members of H-n-plets. These regularities say about approximate invariants of frequencies (or probailities) of members of the alphabets of H-n-plets in long DNA H-sequences. It is obvious that each

of members of alphabets of H-*n*-plets combines a certain set of *n*-plets of nucleotides: for example, the H-doublet 33 combines 4 doublets of nucleotides CC, CG, GC, GG; the H-triplet 332 combines 8 triplets of nucleotides CCA, CCT, CGA, CGT, GCA, GCT, GGA, GGT (see Fig. 22); etc.

　　Fig. 22 shows the genetic matrices from Fig. 4 with a representation of 16 doublets and 64 triplets of nucleotides in forms of appropriate H-doublets and H-triplets.

CC (33)	CA (32)	AC (23)	AA (22)
CT (32)	CG (33)	AT (22)	AG (23)
TC (23)	TA (22)	GC (33)	GA (32)
TT (22)	TG (23)	GT (32)	GG (33)

CCC (333)	CCA (332)	CAC (323)	CAA (322)	ACC (233)	ACA (232)	AAC (223)	AAA (222)
CCT (332)	CCG (333)	CAT (322)	CAG (323)	ACT (232)	ACG (233)	AAT (222)	AAG (223)
CTC (323)	CTA (322)	CGC (333)	CGA (332)	ATC (223)	ATA (222)	AGC (233)	AGA (232)
CTT (322)	CTG (323)	CGT (332)	CGG (333)	ATT (222)	ATG (223)	AGT (232)	AGG (233)
TCC (233)	TCA (232)	TAC (223)	TAA (222)	GCC (333)	GCA (332)	GAC (323)	GAA (322)
TCT (232)	TCG (233)	TAT (222)	TAG (223)	GCT (332)	GCG (333)	GAT (322)	GAG (323)
TTC (223)	TTA (222)	TGC (233)	TGA (232)	GTC (323)	GTA (322)	GGC (333)	GGA (332)
TTT (222)	TTG (223)	TGT (232)	TGG (233)	GTT (322)	GTG (323)	GGT (332)	GGG (333)

Fig. 22. A repetition of genetic matrices from Fig. 4 but with a representation of 16 doublets and 64 triplets of nucleotides in forms of appropriate H-doublets and H-triplets (in brackets).

　　By a comparison of Fig. 22 with Figs. 5, 6 and 17, one can see the following strong correspondence between the genetic unitary matrices (Fig. 6 and 17) and the genetic matrices showing dispositions of H-doublets and H-triplets (Fig. 22): the disposition of each kind of H-doublets and H-tripets in different cells of the matrices in Fig. 22 coincises precisely with the disposition of non-zero entries of different cells of corresponding sparse unitary matrices in Figs. 6 and 17. For example, the disposition of the H-doublet 32 in cells of the matrix of doublets in Fig. 22 coincises with the disposition of non-zero entries in cells of the sparse unitary matrix $R1_4$ in Fig. 6; the disposition of the H-triplet 222 in cells of the matrix of triplets in Fig. 22 coincises with the disposition of non-zero entries in cells of the sparse unitary matrix R_{87} in Fig. 17, etc. Correspondingly, each of the sparse unitary matrices in Figs. 6 and 17 can be defined by dispositions of appropriate H-doublets or H-triplets in cells of matrices in Fig. 22 (taking into account additionally colors of doublets and triplets in Fig. 22 to define signs "plus" or "minus" of entries of these unitary matrices). This connection between the genetic unitary matrices and matrix dispositions of H-doublets and H-triplets in Fig. 22 stimulated us to study frequencies (or probabilities) of H-*n*-plets in long DNA sequences. Results of the study are described below. As we can judge, studies of the frequencies of *n*-plets of hydrogen bonds 3 and 2 in long DNA sequences have never been carried out by anyone early. So, it is a new field for researches.

　　We have studied frequencies of all members of each of H-n-alphabets (for cases *n* = 1, 2, 3, 4, 5) in different H-*n*-plet representations of many long DNA sequences including DNA sequences of complete sets of chromosomes of some organisms. This study has led to interesting evidences in favor of existence of general regularities (principles or laws) concerning frequencies of members of H-*n*-alphabets in long DNA sequences.
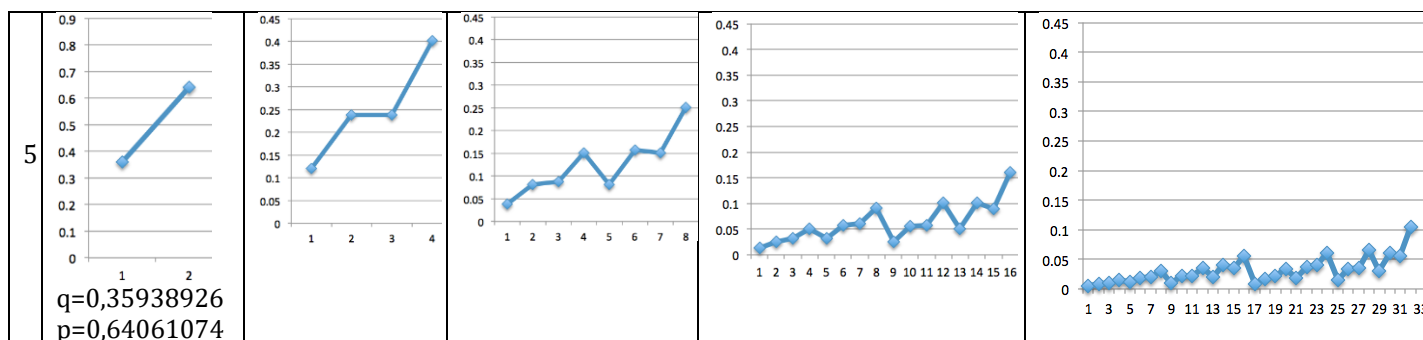
　　As one of our results, Fig. 23 shows frequencies of members of H-n-alphabets (*n* = 1, 2, 3, 4, 5) in DNA sequences of all 5 chromosomes of a plant *Arabidopsis thailana*; initial data about DNA sequences

of these chromosomes were taken from the GenBank: https://www.ncbi.nlm.nih.gov/genome/4. The lengths of these DNA sequences are the following: in the chromosome #1, the length is equal to 30427671 bp; in the chromosome #2 – 19698289; in the chromosome #3 – 23459830; in the chromosome #4 – 18585056; in the chromosome #5 – 26975502.

Members of all these H-n-alphabets in Fig. 23 are given in a certain arrangement corresponding to arrangements of coordinates in the tensor family of vectors $[3, 2]^{(n)}$, where (n) means the tensor power. For example, the vector $[3, 2]^{(2)} = [33, 32, 23, 33]$ have the arrangement of its coordinates 33, 32, 23, 22, which is used in Fig. 23 for the appropriate graph in the case of the alphabet of H-doublets. Similar correspondences hold true for vectors $[3, 2]^{(3)}$, $[3, 2]^{(4)}$, $[3, 2]^{(5)}$ and appropriate graphics of frequencies of members of other H-n-alphabets in Fig. 23.

5	q=0,35938926 p=0,64061074
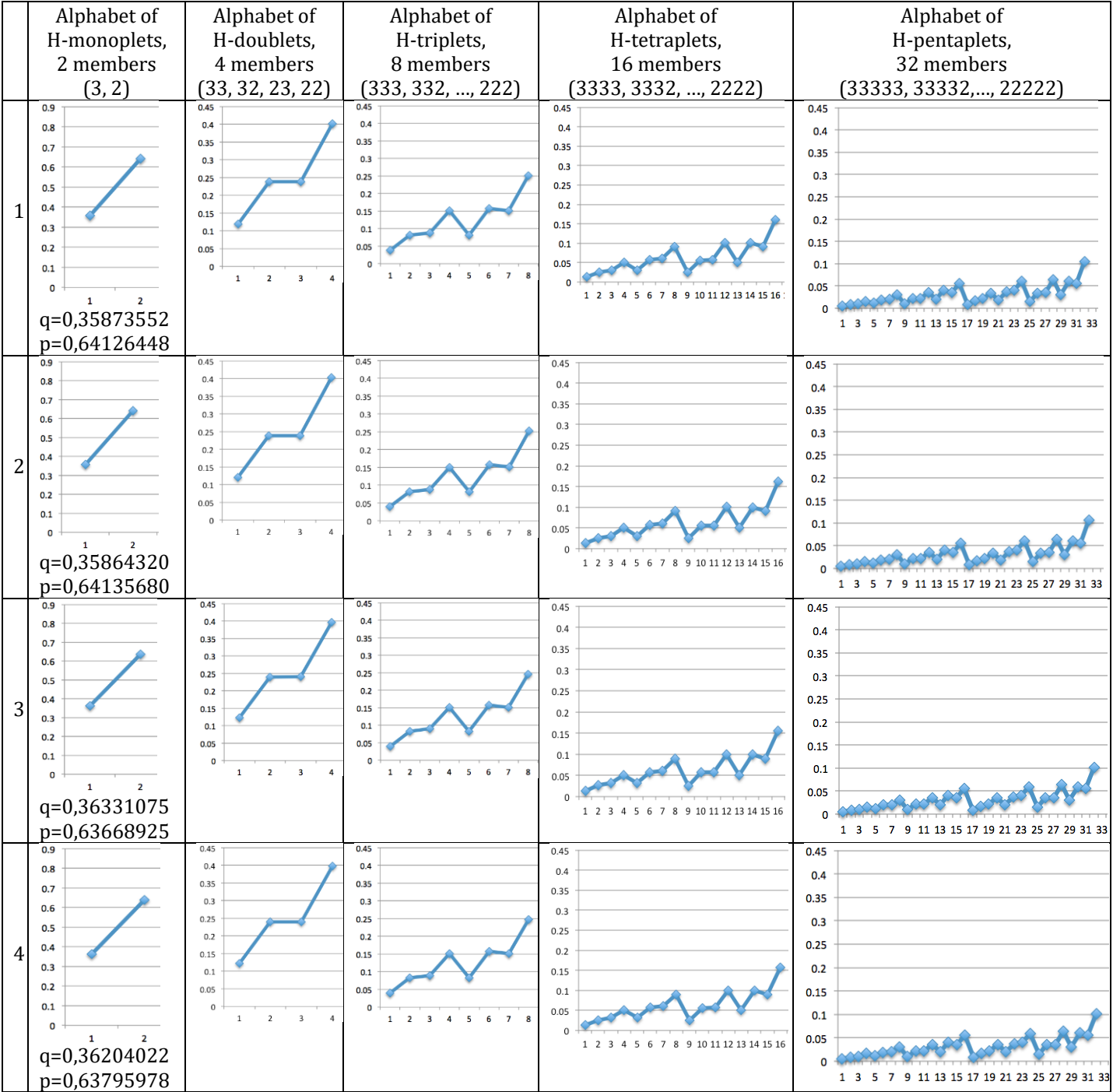
Fig. 23. Frequencies of members of H-*n*-alphabets (*n* = 1, 2, 3, 4, 5) in DNA sequences of all 5 chromosomes of a plant *Arabidopsis thaliana*. The left column contains the sequence numbers of the chromosomes. Each of points in graphs shows a frequency of an appropriate member of H-*n*-alphabets. For each of chromosomes, its frequencies q and p of H-monoplets 3 and 2 correspondingly are shown in the second column. See explanations in the text.

In addition to graphs in Fig. 23, Table 1 shows numeric values of frequencies of all members of these H-*n*-alphabets for all 5 chromosomes and for *n* = 1, 2, 3, 4 (the case *n* = 5 is not shown because its numeric table has a big volume). For statistical analysis of these phenomenolgical data from all chromosomes we use Excel with denotations and formulas from the website on "Real statistics using Excel" (http://www.real-statistics.com/descriptive-statistics/measures-variability/).

Table 1. Frequencies of all members of the H-*n*-alphabets (n = 1, 2, 3, 4) in DNA sequences of all 5 chromosomes of Arabidopsis thaliana are shown. Symbols q and p denote frequencies of H-monoplets 3 and 2 correspondingly and are shown in the left column separately for each chromosome. The frequency values are rounded to the sixth decimal place. Traditional statistical symbols are also shown: $\overline{X}$ denotes the mean of the sample $S = \{x_1, x_2, ..., x_n\}$; s – sample standard deviation; V – coefficient of variation (see their formulas in http://www.real-statistics.com/descriptive-statistics/measures-variability/).

	Alphabet of H-monoplets with 2 members: 3 and 2		Alphabet of H-doublets with 4 members: 33, 32, 23, 22			
	3 (q)	2 (p)	33	32	23	22
Chr. 1	0,358736	0,641264	0,119788	0,239022	0,238874	0,402316
Chr. 2	0,358643	0,641357	0,120564	0,238138	0,238021	0,403277
Chr. 3	0,363311	0,636689	0,122972	0,240226	0,240451	0,396350
Chr. 4	0,362040	0,637960	0,121923	0,240142	0,240093	0,397843
Chr. 5	0,359389	0,640611	0,120390	0,239095	0,238902	0,401612
$\overline{X}$	0,360424	0,639576	0,121127	0,239325	0,239268	0,400280
s	0,001899	0,001899	0,001157	0,000779	0,000886	0,002694
V	0,005270	0,002970	0,009549	0,003254	0,003703	0,006730

Alphabet of H-triplets with 8 members: 333, 332, 323, 322, 233, 232, 223, 222								
	333	332	323	322	233	232	223	222
Chr. 1	0,038514	0,081174	0,088018	0,150704	0,081231	0,157748	0,151366	0,251244
Chr. 2	0,039243	0,081345	0,087551	0,150357	0,081290	0,156850	0,150624	0,252741
Chr. 3	0,040181	0,082637	0,089748	0,150821	0,082845	0,157226	0,150879	0,245662
Chr. 4	0,039645	0,082675	0,089030	0,150815	0,082169	0,157718	0,150903	0,247044
Chr. 5	0,038976	0,081379	0,088099	0,150902	0,081370	0,157821	0,150821	0,250633
$\overline{X}$	0,039312	0,081842	0,088489	0,150720	0,081781	0,157473	0,150919	0,249465
s	0,000569	0,000668	0,000792	0,000192	0,000632	0,000376	0,000244	0,002668
V	0,014480	0,008166	0,008945	0,001273	0,007725	0,002387	0,001619	0,010694

Alphabet of H-tetraplets with 16 members: 3333, 3332, 3323, 3322, 3233, 3232, 3223, 3222, 2333, 2332, 2323, 2322, 2233, 2232, 2223, 2222							
	3333	3332	3323	3322	3233	3232	3223	3222
Chr. 1	0,013226	0,025347	0,030993	0,050243	0,031122	0,056975	0,060129	0,090644
Chr. 2	0,013814	0,025540	0,031157	0,050009	0,031066	0,056623	0,059948	0,090394
Chr. 3	0,013971	0,026168	0,032107	0,050762	0,032179	0,057320	0,061061	0,089600
Chr. 4	0,013625	0,026072	0,031796	0,050559	0,031881	0,057356	0,060887	0,090007
Chr. 5	0,013550	0,025449	0,031281	0,050195	0,031254	0,056887	0,060426	0,090554

$\overline{X}$	0,013637	0,025715	0,031467	0,050354	0,031500	0,057032	0,060490	0,090240
s	0,000253	0,000337	0,000418	0,000270	0,000447	0,000275	0,000427	0,000387
V	0,018529	0,013117	0,013274	0,005366	0,014180	0,004827	0,007059	0,004292
Continuation about frequencies of the second part of members of the alphabet of H-tetraplets:								
	2333	2332	2323	2322	2233	2232	2223	2222
Chr. 1	0,025264	0,056027	0,056957	0,100737	0,050154	0,100824	0,090685	0,160672
Chr. 2	0,025598	0,055669	0,056442	0,100473	0,050130	0,100413	0,090313	0,162411
Chr. 3	0,026132	0,056708	0,057532	0,100124	0,050654	0,100099	0,089708	0,155877
Chr. 4	0,025858	0,056428	0,057387	0,100305	0,050429	0,100297	0,090139	0,156975
Chr. 5	0,025397	0,056027	0,056796	0,100840	0,050101	0,100706	0,090241	0,160293
$\overline{X}$	0,025650	0,056172	0,057023	0,100496	0,050294	0,100468	0,090217	0,159246
s	0,000313	0,000360	0,000396	0,000265	0,000215	0,000265	0,000314	0,002435
V	0,012221	0,006409	0,006945	0,002639	0,004279	0,002642	0,003483	0,015294

One can see from Table 1, that the frequency of each of members of these H-n-alphabets has approximately the same value in all 5 chromosomes or, in other words, it serves as a special invariant inside the complete set of chromosomes *Arabidopsis thaliana*. It is obvious that these reqularities of frequencies of the hydrogen bonds 3 and 2 and of their ensembles define in long DNA sequences a special limitation for arrangements of nucleotides A, T, C, G, which are connected with hydrogen bonds in DNA sequences (A=T=2, C=G=3): only special classes of locations of the nucleoitides in a long sequence can provide such regularities.

The data from Table 1 that all 5 chromosomes have approximately the identical value q of frequencies of the H-monoplet 3 and also approximately the identical value p of frequencies of the H-monoplet 2 remind the following known data: the article [Okamura, Wei, Scherer, 2007] shows an approximate equality of frequencies of each of nucleotides A, T, C and G in all human nuclear chromosomes with a precision in a few percentage. Correspondingly a summary frequency of nucleotides A and T (and also C and G), which is directly connected with the frequency of the H-monoplet 2 (the H-monoplet 3), has approximately the same values in all human nuclear chromosomes.

Similar results (with some other levels of accuracy) have been obtained in our analysis of the following complete sets of nuclear chromosomes on the basis of initial data about their DNA sequences from the GenBank: a nematode *Caenorhabditis elegans*, fruit fly *Drosophila melanogaster*, house mouse *Mus musculus*, Homo sapiens.

These results allow assuming existence of the following general rule for eucariots about frequencies of hydrogen bonds in complete sets of nuclear chromosomes:

- in complete sets of nuclear chromosomes of different organisms, a frequency of any member of the H-n-alphabets has approximately the same value in all chromosomes.

Of course, further researches are needed to define a degree of universality and precision of this rule. This rule has mathematical analogies with the main law of population genetics - the Hardy–Weinberg law as it is described in the next Section.

7. The mathematical model of frequencies of hydrogen bonds in long DNA sequences

Our conception of the quantum-algorithmic genetics (or the quantum-computing genetics) [Petoukhov, 2017a; Petoukhov, Svirin, 2018] has led the author to the useful mathematical model not only for phenomenologic data presented in Fig. 23 but also for similar data in many other cases of long DNA sequences studied in our researches. In the end of this Section we describe this model using the notion of qubits and also some classical formalisms from quantum informatics [Nielsen, Chuang, 2010]. But firstly we describe our model approach by means of a simpler language.

In our proposed approach, phenomenologic frequencies of all members of H-n-alphabets $[3, 2]^{(n)}$ in a long DNA sequence under fixed n are modeled by numeric values of appropriate coordinates of 2^n-dimensional vector from the tensor family of frequency vectors $[q, p]^{(n)}$ where q and p mean frequencies of the H-monoplet 3 and the H-monoplet 2 correspondingly (these frequencies q and p are shown for each of chromosomes in the second columns in Fig. 23). Correspondingly, in this model approach, to get model values of frequencies of all 2^n members of the H-n-alphabet in a considered long DNA, one should calculate the following 2^n coordinates of frequency vectors from their tensor family:

- in the case of H-doublets-alphabet, 4 coordinates of the vector $[q, p]^{(2)} = [qq, qp, pq, pp]$;

- in the case of H-triplets-alphabet, 8 coordinates of the vector [*q, p*]$^{(3)}$ = [*qqq, qqp, qpq, qpp, pqq, pqp, ppq, ppp*];
- in the case of H-tetraplets-alphabet, 16 coordinates of the vector [*q, p*]$^{(4)}$ = [*qqqq, qqqp, qqpq, qqpp, qpqq, qpqp, qppq, qppp, pqqq, pqqp, pqpq, pqpp, ppqq, ppqp, pppq, pppp*];
- in the case of H-pentaplets-alphabet, 32 coordinates of the vector [*q, p*]$^{(5)}$ = [*qqqqq, qqqqp, ... , ppppp*], etc.

Briefly speaking, in line with the proposed model approach, for a prediction of approximate values of frequencies «f» of many members of H-n-alphabets in a long DNA sequence on the basis of knowledge about only two frequencies *q* and *p* , it is enough to do the following:

- calculate frequencies *q* and *p* of H-monoplets 3 and 2 in the sequence;
- replace digits 3 and 2 in H-symbols of these H-alphabetic members by their frequencies *q* and *p* as factors;
- calculate the product of these factors.

The expression (14) shows an example of such calculation of the frequency f(332) for the member 332 of the H-triplets-alphabet under *q*=0,36 and p=0,64 (*q*+*p*=1):

$$f(332) = qqp = 0,36*0,36*0,64 = 0,0829 \qquad (14)$$

Fig. 24 shows nice correspondences between these model values (red points) of components of frequency vectors [*q, p*]$^{(n)}$ and phenomenological values (blue points) of frequencies of all members of H-*n*-alphabets (*n* = 2, 3, 4, 5). One can see in these graphs that the model points of red color turned out to be almost exactly superimposed on the phenomenological points of blue color. It gives evidences in favor that knowing only two probablities *q* and *p* of H-monoplets 3 and 2 in a long DNA sequence, one can predict dozens of frequency values of all members of H-*n*-alphabets for this DNA sequence with a qood level of accuracy. We presume that a similar model correspondence holds true also for *n* = 6, 7, 8, ... (if *n* is not too large) but this should be studied in future researches.

	Alphabet of H-monoplets, 2 members (3, 2)	Alphabet of H-doublets, 4 members (33, 32, 23, 22)	Alphabet of H-triplets, 8 members (333, 332, ..., 222)	Alphabet of H-tetraplets, 16 members (3333, 3332, ..., 2222)	Alphabet of H-pentaplets, 32 members (33333, 33332,..., 22222)
1	*q*=0,35873552 *p*=0,64126448	[*q, p*]$^{(2)}$	[*q, p*]$^{(3)}$	[*q, p*]$^{(4)}$	[*q, p*]$^{(5)}$
2	*q*=0,35864320 *p*=0,64135680	[*q, p*]$^{(2)}$	[*q, p*]$^{(3)}$	[*q, p*]$^{(4)}$	[*q, p*]$^{(5)}$
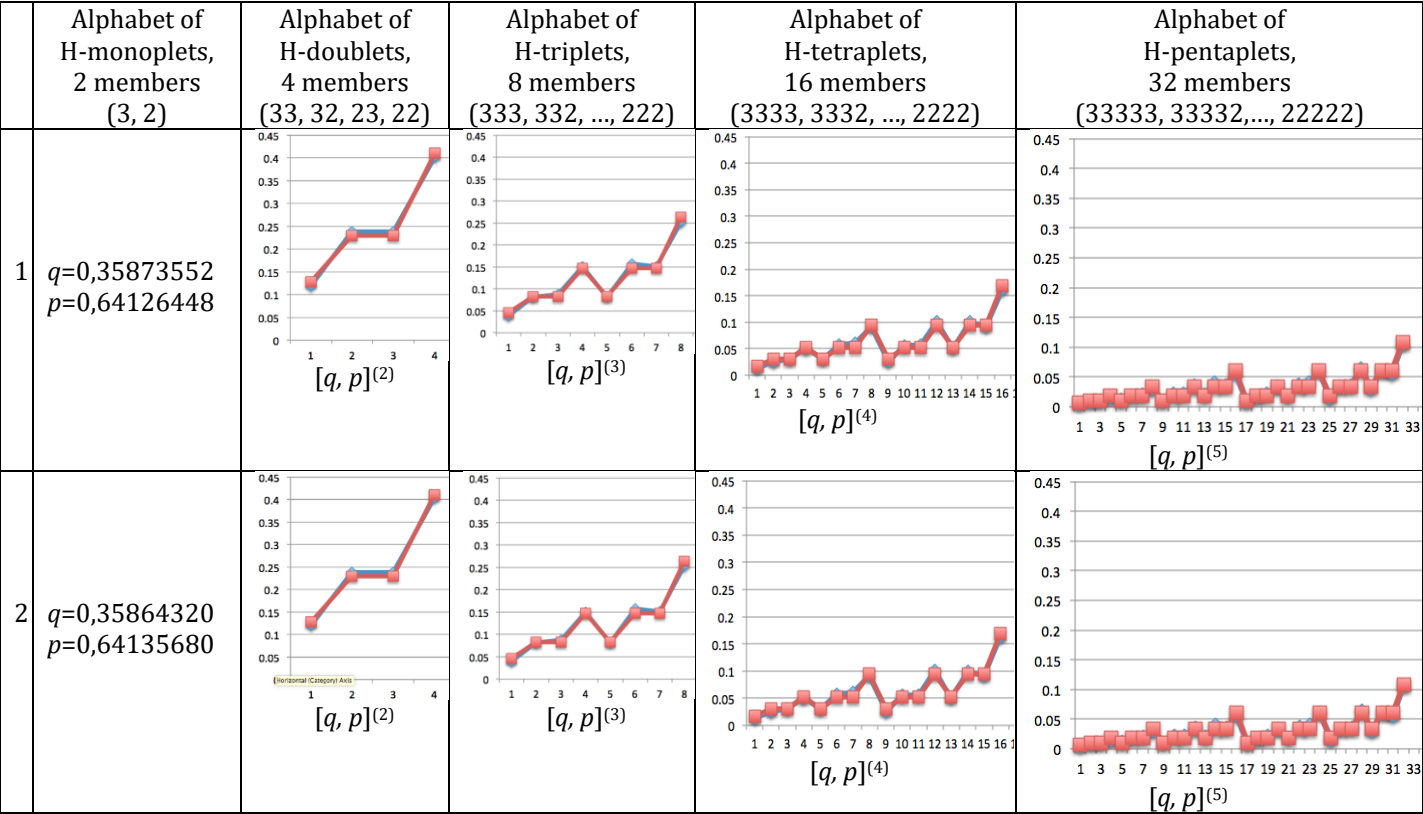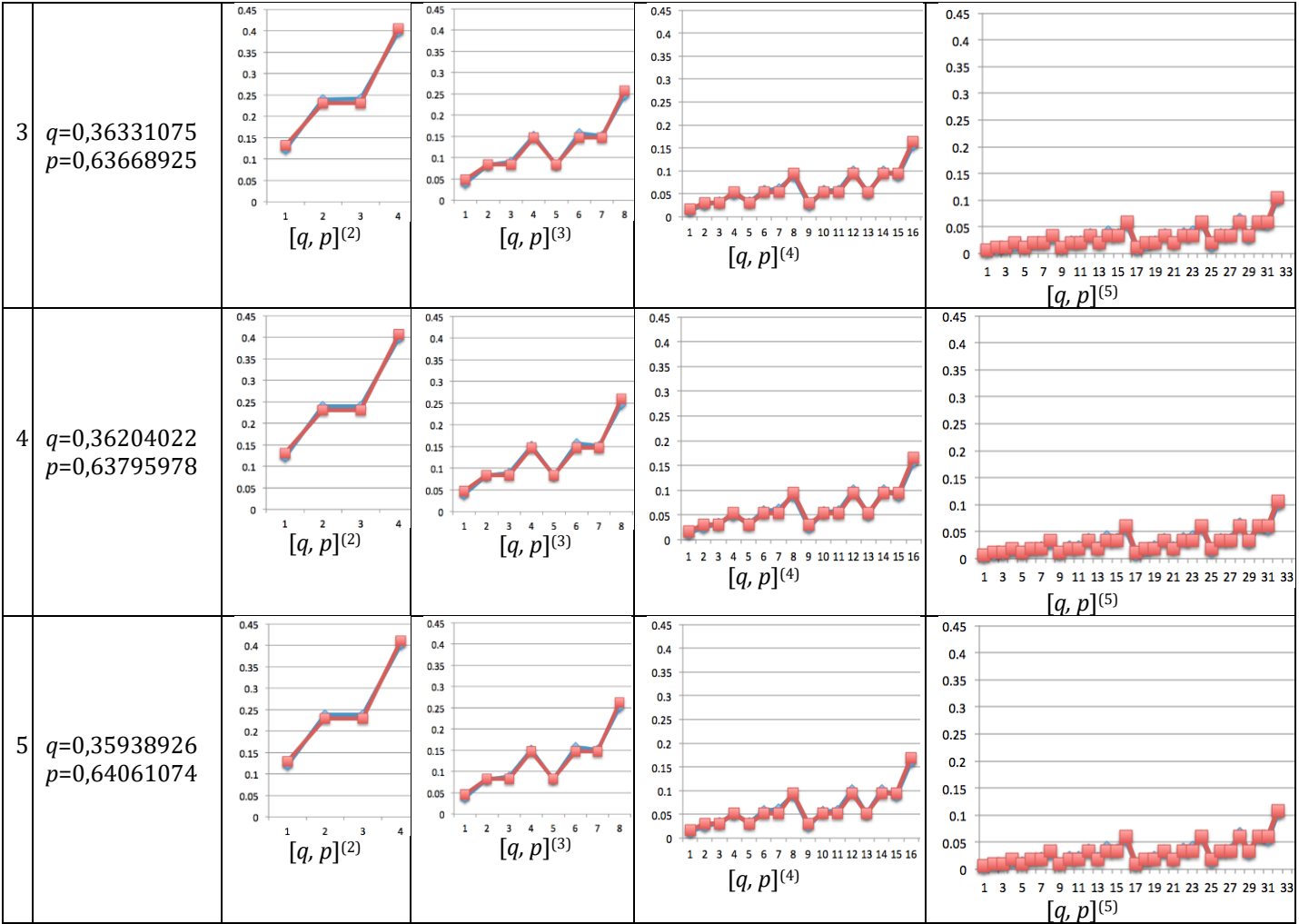
Fig. 24. The nice correspondence between coordinate values (red points) of model vectors $[q, p]^{(n)}$ and phenomenological values (blue points taken from Fig. 23) of frequencies of all members of H-*n*-alphabets (*n* = 2, 3, 4, 5) for DNA sequences of all 5 chromosomes of a plant *Arabidopsis thaliana*. The left column contains the sequence numbers of the chromosomes. See explanations in the text.

In addition to graphs in Fig. 24, Table 2 shows numeric values of coordinates of vectors $[q, p]^{(n)}$ for all 5 chromosomes and for *n* = 1, 2, 3, 4 (the case *n* = 5 is not shown because its numeric table needs a big place and can be created by analogy). By a comparison of Tables 1 and 2, one can see that the proposed model approach gives nice results in modeling the phenomenological frequencies.

Table 2. Values of coordinates of vectors $[q, p]^{(n)}$ are shown, which are used as models of values of frequencies of all members of the H-*n*-alphabets (*n* = 1, 2, 3, 4) in DNA sequences of all 5 chromosomes of *Arabidopsis thaliana*. Symbols *q* and *p* denote frequencies of H-monoplets 3 and 2 correspondingly and are shown in the left column separately for each chromosome. The model values are rounded to the sixth decimal place. Traditional statistical symbols are also shown: $\overline{X}$ denotes the mean of the sample $S = \{x_1, x_2, ..., x_n\}$; s – sample standard deviation; V – coefficient of variation (see their formulas in http://www.real-statistics.com/descriptive-statistics/measures-variability/).

	Alphabet of H-monoplets with 2 members (3, 2)		Alphabet of H-doublets with 4 members (33, 32, 23, 22)			
	MODEL VALUES:					
	3 (q)	2 (p)	33 (qq)	32 (qp)	23 (pq)	22 (pp)
Chr. 1	0,358736	0,641264	0,128691	0,230044	0,230044	0,411220
Chr. 2	0,358643	0,641357	0,128625	0,230018	0,230018	0,411339
Chr. 3	0,363311	0,636689	0,131995	0,231316	0,231316	0,405373
Chr. 4	0,362040	0,637960	0,131073	0,230967	0,230967	0,406993
Chr. 5	0,359389	0,640611	0,129161	0,230229	0,230229	0,410382

X̄	0,360424	0,639576	0,129909	0,230515	0,230515	0,409061
s	0,001899	0,001899	0,001371	0,000528	0,000528	0,002428
V	0,005270	0,002970	0,010553	0,002293	0,002293	0,005935

Alphabet of H-triplets with 8 members (333, 332, 323, 322, 233, 232, 223, 222)								
	333	332	323	322	233	232	223	222
MODEL VALUES:								
	ααα	ααβ	αβα	αββ	βαα	βαβ	ββα	βββ
Chr. 1	0,046166	0,082525	0,082525	0,147519	0,082525	0,147519	0,147519	0,263701
Chr. 2	0,046130	0,082494	0,082494	0,147524	0,082494	0,147524	0,147524	0,263815
Chr. 3	0,047955	0,084040	0,084040	0,147276	0,084040	0,147276	0,147276	0,258097
Chr. 4	0,047454	0,083619	0,083619	0,147348	0,083619	0,147348	0,147348	0,259645
Chr. 5	0,046419	0,082742	0,082742	0,147487	0,082742	0,147487	0,147487	0,262895
X̄	0,046825	0,083084	0,083084	0,147431	0,083084	0,147431	0,147431	0,261631
s	0,000742	0,000629	0,000629	0,000100	0,000629	0,000100	0,000100	0,002328
V	0,015849	0,007568	0,007568	0,000681	0,007568	0,000681	0,000681	0,008896

Alphabet of H-tetraplets with 16 members (3333, 3332, 3323, 3322, 3233, 3232, 3223, 3222, 2333, 2332, 2323, 2322, 2233, 2232, 2223, 2222)								
Model values for the first half of members of the alphabet of H-tetraplets:								
	3333	3332	3323	3322	3233	3232	3223	3222
	qqqq	qqqp	qqpq	qqpp	qpqq	qpqp	qppq	qppp
Chr. 1	0,016561	0,029605	0,029605	0,052920	0,029605	0,052920	0,052920	0,094599
Chr. 2	0,016544	0,029586	0,029586	0,052908	0,029586	0,052908	0,052908	0,094615
Chr. 3	0,017423	0,030532	0,030532	0,053507	0,030532	0,053507	0,053507	0,093769
Chr. 4	0,017180	0,030274	0,030274	0,053346	0,030274	0,053346	0,053346	0,094002
Chr. 5	0,016682	0,029736	0,029736	0,053005	0,029736	0,053005	0,053005	0,094482
X̄	0,016878	0,029947	0,029947	0,053137	0,029947	0,053137	0,053137	0,094293
s	0,000357	0,000385	0,000385	0,000244	0,000385	0,000244	0,000244	0,000344
V	0,021158	0,012857	0,012857	0,004588	0,012857	0,004588	0,004588	0,003649
Model values for the second part of members of the alphabet of H-tetraplets:								
	2333	2332	2323	2322	2233	2232	2223	2222
	pqqq	pqqp	pqpq	pqpp	ppqq	ppqp	pppq	pppp
Chr. 1	0,029605	0,052920	0,052920	0,094599	0,052920	0,094599	0,094599	0,169102
Chr. 2	0,029586	0,052908	0,052908	0,094615	0,052908	0,094615	0,094615	0,169199
Chr. 3	0,030532	0,053507	0,053507	0,093769	0,053507	0,093769	0,093769	0,164327
Chr. 4	0,030274	0,053346	0,053346	0,094002	0,053346	0,094002	0,094002	0,165643
Chr. 5	0,029736	0,053005	0,053005	0,094482	0,053005	0,094482	0,094482	0,168413
X̄	0,029947	0,053137	0,053137	0,094293	0,053137	0,094293	0,094293	0,167337
s	0,000385	0,000244	0,000244	0,000344	0,000244	0,000344	0,000344	0,001983
V	0,012857	0,004588	0,004588	0,003649	0,004588	0,003649	0,003649	0,011853

We have calculated also frequencies of members of H-*n*-alphabets ($n = 1, 2, 3, 4, 5$) in 19 genomes of bacteria and archaea from the full list in the article [Rapoport, Trifonov, 2012, p. 2]: "*Aquifex aeolicus, Acidobacteria bacterium, Bradyrhizobium japonicum, Bacillus subtilis, Chlamydia trachomatis, Chromobacterium violaceum, Dehalococcoides ethenogenes, Escherichia coli, Flavobacterium psychrophilum, Gloeobacter violaceus, Helicobacter pilory, Methanosarcina acetivorans, Nanoarchaeum equitans, Syntrophus aciditrophicus, Streptomyces coelicolor, Sulfolobus solfataricus, Treponema denticola, Thermotoga maritima and Thermus thermophiles*". The calculated sets of frequencies were also modelled on the basis of sets of coordinates of appropriate vectors $[q, p]^{(n)}$ with good results (with different levels of accuracy in different genomes). These results confirm that the proposed model approach on the basis of vectors $[q, p]^{(n)}$ can be used to get idealized models of frequencies of the members of H-*n*-alphabets in real long DNA sequences. These received data of the analysis for genomes of prokaryots will be published some later.

Let us analyse the tensor family of vectors $[q, p]^{(n)}$ in more details. One can see that in these vectors some values of coordinates, which indicate appropriate frequencies, are repeated. For example, the set of coordinates of the vector $[q, p]^{(2)} = [qq, qp, pq, pp]$ has 1 value qq, 2 values qp (=pq), 1 value pp; the set of coordinates of the vector $[q, p]^{(3)} = [qqq, qqp, qpq, qpp, pqq, pqp, ppq, ppp]$ has 1 value qqq, 3 values qqp (=qpq=pqq), 3 values qpp (=pqp=ppq) and 1 value ppp; etc. These numbers of separate frequencies as coordinates of vectors $[q, p]^{(n)}$ are formed the Pascal's triangle (https://en.wikipedia.org/wiki/Pascal%27s_triangle) connected with the Newton's binomial formula

(https://www.lucaswillems.com/en/articles/5/newtons-binomial-formula). This Pascal's triangle is shown in Fig. 25.

	Amounts of identical values of coordinates in vectors $[q, p]^{(n)}$									
n=1				1		1				
n=2			1		2		1			
n=3		1		3		3		1		
n=4	1		4		6		4		1	
n=5	1	5		10		10		5		1
...	.	.	.	.	.	.	.	.	.	.

Fig. 25. The Pascal's triangle for amounts of identical values of coordinates in vectors of the tensor family $[q, p]^{(n)}$.

But applications of the Newton's binomial formula in Mendelian genetics and in population genetics are known for a long time, first of all, in a connection with the Hardy-Weinberg law (https://en.wikipedia.org/wiki/Hardy%E2%80%93Weinberg_principle; https://en.wikipedia.org/wiki/Allele_frequency ; Crow, 1999; Gillespie, 2004). «*The Hardy-Weinberg law is the cornerstone of diploid population genetics. Yet it seems trivially obvious, a routine application of the binomial theorem*» [Crow, 1999, p. 821]. This law is also called as the basis of mathematical constructions in population genetics and contemporary evolutionary theory. The Hardy-Weinberg law known from 1908 year states that allele and genotype frequencies in a population will remain constant from generation to generation in the absence of other evolutionary influences. «*These influences include genetic drift, mate choice, assorttative mating, natural selection, sexual selection, mutation, gene flow, meiotic drive, genetic hitchhiking, population bottleneck, founder effect and inbreeding*» (https://en.wikipedia.org/wiki/Hardy%E2%80%93Weinberg_principle). By these reasons, the Hardy-Weinberg law is not a precise but idealized model law in a general case.

This law uses the Hardy-Weinberg equation (15) to examine a simple genetic locus at which there are two alleles, A and a:

$$q^2 + 2qp + p^2 = 1 \tag{15}$$

where q is the frequency of the «a» allele and p is the frequency of the «A» allele in the population. In the equation (15), q^2 represents the frequency of the homozyqous genotype aa, p^2 represents the frequency of the homozyqous genotype AA, and 2qp represents the frequency of the heterozyqous genotype Aa. The sum of the allele frequencies for all the alleles at the locus must be 1, so p + q = 1. If the p and q allele frequencies are known, then the frequencies of the three genotypes may be calculated using the Hardy-Weinberg equation (https://www.nature.com/scitable/definition/hardy-weinberg-equation-299). In population genetics studies, the Hardy-Weinberg equation (14) is used in particularly to measure whether the observed genotype frequencies in a population differ from the frequencies predicted by the equation.

Now we should describe that in our analysis of the tensor families of H-*n*-alphabets $[3, 2]^{(n)}$ and frequency vectors $[q, p]^{(n)}$, we have got in some sense the similar equation (16) and the similar rule for frequencies of members of H-*n*-alphabets though the obvious great difference exists between genotype frequencies in population genetics and frequencies of members of H-*n*-alphabets in long DNA sequences. As it was mentioned above, triplets with strong and weak roots have different properties concerning the degeracy of the genetic code (Fig. 4). We will also distinguish in each of the H-*n*-alphabets $[3, 2]^{(n)}$ its members on the basis of their roots, that is, on the basis of their first two H-monoplets:

-   members with a root composed of identical H-monoplets will be called as homo-rooted members of the H-*n*-alphabet by analogy with the term "homozygous genotype" (for example, the H-tetraplets **33**23, **33**22, **33**32, **33**33 belong to homo-rooted members relative to   H-

monoplet 3; the H-tetraplets **22**33, **22**32, **22**23, **22**22 belong to homo-rooted members relative to H-monoplet 2);

- members with a root composed of different H-monoplets will be called as hetero-rooted members of the H-*n*-alphabet by analogy with the term "heterozygous genotype" (for example, the H-tetraplets **32**33, **32**22, **23**33, **23**22 belong to hetero-rooted members of the   H-4-alphabet represented by components of the vector [3, 2]$^{(4)}$).

Let us calculate for each of H-*n*-alphabetic vectors [3, 2]$^{(n)}$ (*n* = 2, 3, 4, ...) the sums of frequencies of its homo-rooted members and separately the sum of its hetero-rooted members, taking into account that frequencies of all these members are represented by appropriate components of the frequency vectors [q, p]$^{(n)}$ where *q+p*=1. Here *q* means a frequency of the H-monoplet 3, *p* means the frequency of the H-monoplet 2. The calculations give the following results.

In the case of the H-doublets-alphabet, its 4 members are components of the vector [3, 2]$^{(n)}$ = [33, 32, 23, 22] and their frequencies are represented by appropriate components of the frequency vector [q, p]$^{(2)}$ = [*qq, qp, pq, pp*]. Correspondingly the frequency of the homo-rooted member 33 is equal to q^2; the frequency of the homo-rooted member 22 is equal to p^2 and the sum of frequencies of the hetero-rooted members 32 and 23 is equal to 2*qp*. The total sum of all these frequencies should be equal to 1. This is expressed by the equation (15), which is completly analogical to the Hardy-Weinberg equation (14):

$$q^2 + 2qp + p^2 = 1 \qquad\qquad\qquad (16)$$

In the case of the H-triplets-alphabet, its 8 members are components of the vector [3, 2]$^{(3)}$ = [333, 332, 323, 322, 233, 232, 223, 222] and their frequencies are represented by appropriate components of the frequency vector [q, p]$^{(3)}$ = [*qqq, qqp, qpq, qpp, pqq, pqp, ppq, ppp*]. In this case, the sum of frequencies of homo-rooted members 333 and 332 is equal to q^2 again since (*qqq + qqp)* = *qq**(*q+p)* = q^2; the sum of frequencies of homo-rooted members 223 and 222 is equal to p^2 again since (*ppq + ppp*) = *pp*(*q+p*) = p^2 and the sum of frequencies of all hetero-rooted members 323, 322, 233, 232 is equal to 2*qp* again since (*qpq + qpp + pqq + pqp*) = *qp*(*q+p*) +*pq*(*q+p*) = 2*qp* (taking into account that *q+p*=1). So, in the case of the H-triplets-alphabet the equation (16) is also true.

In the case of H-tetraplets-alphabet, its 16 members are components of the vector [3, 2]$^{(4)}$ = [3333, 3332, 3323, 3322, 3233, 3232, 3223, 3222, 2333, 2332, 2323, 2322, 2233, 2232, 2223, 2222] and their frequencies are represented by appropriate components of the frequency vector [q, p]$^{(4)}$ = [*qqqq, qqqp, qqpq, qqpp, qpqq, qpqp, qppq, qppp, pqqq, pqqp, pqpq, pqpp, ppqq, ppqp, pppq, pppp*]. In this case, the sum of frequencies of homo-rooted members 3333, 3332, 3323, 3322 is equal to q^2 again since (*qqqq + qqqp + qqpq + qqpp*) = *qqq*(*q+p*) +*qqp*(*q+p*) = *qqq* + *qqp* = *qq*(*q+p*) = q^2; the sum of frequencies of homo-rooted members 2233, 2232, 2223, 2222 is equal to p^2 again since (*ppqq + ppqp + pppq + pppp*) = *ppq*(*q+p*) + *ppp*(*q+p*) = *ppq* +*ppp* = *pp*(*q+p*) = p^2; the sum of frequencies of all hetero-rooted members 3233, 3232, 3223, 3222, 2333, 2332, 2323, 2322 is equal to 2*qp* again since (*qpqq + qpqp + qppq + qppp + pqqq + pqqp + pqpq + pqpp*) = *qpq*(*q+p*) + *qpp*(*q+p*) + *pqq*(*q+p*) + *pqp*(*q+p*) = *qpq* + *qpp* + *pqq* + *pqp* = *qp*(*q+p*) + *pq*(*q+p*) = 2*qp*. So, in the case of the H-tetraplets-alphabet the equation (16) is also true.

By analogy one can check that in other cases of H-*n*-alphabets with *n* = 5, 6, 7, ... the equation (16) are also true. For all these cases, in the equation (16) the summand q^2 means the total frequencies of all homo-rooted members relative to the H-monoplet 3; the summand p^2 means the total frequencies of all homo-rooted members relative to the H-monoplet 2; the summand 2*qp* means the total frequencies of all hetero-rooted members in any H-*n*-alphabet (*n* = 2, 3, 4, ...). On this basis, we propose the following general and idealized rule in the frame of our model approach (by analogy with the formulation of the Hardy-Weinberg law):

- Frequencies of homo-rooted and hetero-rooted members of H-*n*-alphabets (*n* = 2, 3, ,4, ...) in any long DNA sequence will remain constant from one H-*n*-alphabet to another H-*n*-alphabet in the absence of special evolutionary influences.

But what one can say about these "special evolutionary influences"? Briefly speaking from the standpoint of quantum informatics (see the last part of this Section), these are those influences, which disturb so called separable pure states of long DNA sequences and transform them into entangled states. As known, "*the state space of a composite physical systems is the tensor product of the state spaces of the component physical systems. Moreover, if we have systems numbered 1 through n and system*

number i is prepared in the state ρ_i, then the joint state of the total system is $\rho_1 \otimes \rho_2 \otimes ... \rho_n$" [Nielsen, Chuang, 2010, p. 102]. If a quantum state can be represented as a vector of a Hilbert space H_i, such state is called a pure quantum state. If a pure state $|\psi> \in H_1 \otimes H_2$ can be written in the form $|\psi> = |\psi_1> \otimes |\psi_2>$, where $|\psi_i>$ is a pure state of the i-th subsystem, it is said to be separable. Otherwise it is called entangled. In our model approach, we represent long DNA sequences as quantum systems in separable pure states. Any discrepancies between real and model values of frequencies (or probabilities) of members of H-n-alphabets in long DNA sequences can be attributed to the violation of separable pure states of the considered model systems due to their transition to entangled states under different evolutionary influences such as mutations, dreifs of genes, etc.

We believe that expressions of the sort (14) and the equation (16) could play a significant role for structural analysis of long DNA sequences and their evolution by some analogy with the important role of the Hardy-Weinberg equation (15) in population genetics and evolutionary biology. Of course, a degree of universality of this rule about frequencies of members of H-n-alphabets in long DNA sequences should be tested and defined in future researches. By some analogy with the Hardy-Weinberg equation (15), expressions of the sort (14) and the equation (16) can be used in particularly to measure whether the observed frequencies of members of H-n-alphabets in long DNA sequences differ from frequencies predicted by them. One can suppose that genetic phenomena described by the Hardy-Weinberg law have an evolutionary connection with the described phenomena of long DNA sequences of hydrogen bonds 3 and 2.

We obtained the same results initially on the basis of applications of notions and formalisms of quantum compiters to analysis of long DNA sequences. First of all, we use the notion of n-qubit systems and the algebraic operation of the tensor product of matrices and vectors. As we already noted above, the tensor product is the crucial operation to understanding the quantum mechanics of multiparticle systems [Nielsen, Chuang, 2010, p. 71] and is one of basic instruments in quantum informatics.

As known, a quantum bit (or qubit) is a unit of quantim information. For two-level quantum systems used as qubits, the state $|0>$ is identified with the vector $(1, 0)$, and similarly the state $|1>$ with the vector $(0, 1)$. Two possible states for a qubit are the states $|0>$ and $|1>$, which correspond to the states of 0 and 1 for a classical bit. The difference between bits and qubits is that a qubit can be in a state other than $|0>$ or $|1>$. It is possible to form linear combinations of states, often called superpositions (17):

$$|\psi> = \sqrt[2]{q} \, |0> + \sqrt[2]{p} \, |1>. \tag{17}$$

Here the symbol $|\psi>$ means a state of a gubit. The numbers $\sqrt[2]{q}$ and $\sqrt[2]{p}$ can be complex numbers but in our case it is enough to think of them as real numbers. Put another way, the state of a qubit is a vector in a two-dimensional vector space. The standart notation for states in quantum mechanics is the Dirac notation "$| >$". The special states $|0>$ and $|1>$ are known as computational basis states, and form an orthonormal basis for this vector space [Nielsen, Chuang, 2010, p. 13]. As known, we cannot examine a qubit to determine its quantum state, that is, the values of q and p. Instead, quantum mechanics tells us that we can only acquire much more restricted information about the quantum state. When we measure a qubit we get either the result 0, with probability q, or the result 1, with probability p. Naturally, the probabilities must sum to one (18):

$$q + p = 1 \tag{18}$$

Geometrically, we can interpret this as the condition that the qubit's state be normalized to length 1. Values $\sqrt[2]{q}$ and $\sqrt[2]{p}$ are called amplitudes of probabilities q and p (the terms of "probabilities" and "frequencies" are synonyms in this article). Thus, in general a qubit's state is a unit vector in a two-dimensional complex vector space. Let us emphasize again that when a qubit is measured, it only ever gives "0" or "1" as the measurement result – probabilistically.

In more general case, a system of "n" qubits is considered in quantum informatics. The computational basis states of this system are written in the form $|x_1 x_2 .... x_n>$; a quantum state of such a system is specified by 2^n amplitudes [Nielsen, Chuang, 2010, p. 17]. In our model approach we interpret different H-*n*-representations (described below) of DNA texts as quantum systems of   *n*-qubits.

In technical devices of quantum informatics, a qubit can be represented by many ways on the basis of different pairs of binary-oppositional indicators: for example, by two electronic levels in an atom; by two kinds of polarization of a single photon (vertical polarization and horizontal polarization), etc.

In our model approach for genetic informatics, we represent a qubit as a two-level quantum system: in this quantum system of a long DNA sequence one level corresponds to the indicator   «3 hydrogen bonds» and the second level – to the oppositional indicator «2 hydrogen bonds». In other words, such «genetic» qubit is represented by these oppositional indicators and the state of such qubit is a vector in its appropriate two-dimensional Hilbert space H. One can assume that the state |0> corresponds to the state «3 hydrogen bonds», and the state |1> - to the state «2 hydrogen bonds». In this case, the state of a long DNA sequence as a quantum system in the form of a sequence of numbers 3 and 2 of hydrogen bonds is described by the expression (14) where q and p are probabilities (or frequencies) of its nucleotides with 3 or 2 hydrogen bonds correspondingly.

The tensor product of the two-dimensional Hilbert space H⊗H gives one 4-dimensional Hilbert space with the following separable pure state of a quantum 2-qubit system (19):

$$|\psi>^{(2)} = |\psi> \otimes |\psi> = (\sqrt[2]{q}\,|0> + \sqrt[2]{p}\,|1>) \otimes (\sqrt[2]{q}\,|0> + \sqrt[2]{p}\,|1>) =$$
$$(\sqrt[2]{q} * \sqrt[2]{q}\,)|00> + (\sqrt[2]{q} * \sqrt[2]{p})|01> + (\sqrt[2]{p} * \sqrt[2]{q}\,)|10> + (\sqrt[2]{p} * \sqrt[2]{p})|11> \qquad (19)$$

Such 2-qubit system (~~16~~ 19), which is generated by the second tensor power $|\psi>^{(2)}$ of a state $|\psi>$ of a qubit (14), has 4 computational basis states denoted |00>, |01>, |10>, |11>. Probabilities (or frequencies) of these states satisfy the normalization condition (20):

$$(\sqrt[2]{q} * \sqrt[2]{q}\,)^2 + (\sqrt[2]{q} * \sqrt[2]{p})^2 + (\sqrt[2]{p} * \sqrt[2]{q}\,)^2 + (\sqrt[2]{p} * \sqrt[2]{p})^2 = qq + qp + pq + pp = 1. \qquad (20)$$

One can see the following strong correspondence between the denotations of computational basis states |00>, |01>, |10>, |11> and appropriate values of probabilities for them qq, qp, pq, pp: in any of the denotations of these computational basis states, replacements of the symbols 0 and 1 by values q and p correspondingly give the value of the probability of this state. For example, such replacement in the denotation of the state |01> gives its probability qp. Such strong correspondence between denotations of computational basis states and values of their probabilities as the product of the probabilities q and p holds true in other cases of the tensor powers $|\psi>^{(n)}$ of a state $|\psi>$ of a qubit (17).

The third tensor power $|\psi>^{(3)}$ of a state $|\psi>$ of a qubit (14) gives a 3-qubit system with 8 computational basis states denoted |000>, |001>, |010>, |011>, |100>, |101>, |110>, |111>. One can check that denotations of these states |000>, |001>, |010>, |011>, |100>, |101>, |110>, |111> coincide with binary-numeric denotations of appropriate triplets in cells of the genetic (8*8)-matrix in Fig. 4. Probabilities of these states are qqq, qqp, qpq, qpp, pqq, pqp, ppq, ppp correspondingly and they satisfy the normalization condition (18):

$$qqq + qqp + qpq + qpp + pqq + pqp + ppq + ppp = 1. \qquad (21)$$

The nth tensor power $|\psi>^{(n)}$ of a state $|\psi>$ of a qubit (17) gives an appropriate n-qubit system with 2^n computational basis states, each of which has its own value of a probability (or a frequency) calculated by analogy; formally it is possible simply replace symbols 0 and 1 in the denotations of the computational basis states by values q and p correspondingly to get values of probabilities of members of the H-n-alphabet in this model of a long DNA text.

8. Unitary genetic matrices, frequencies of hydrogen bonds and hypercomplex numbers

Above we have described some evidences in favor that structurisation of long DNA texts is connected with frequencies of members of H-n-alphabets in the texts. Now let us pay our attention initially to the H-doublets-alphabet with its 4 members 33, 32, 23, 22. Frequencies of these members in a real DNA sequence will be denoted by symbols α_0, α_1, α_2, and α_3 correspondingly. Fig. 22 shows locations of the members in the genetic (4*4)-matrix of 16 doublets. If in this matrix the members 33,

32, 23, 22 are replaced by their frequencies α_0, α_1, α_2, and α_3, the numeric matrix F_4 of the frequencies arises. Fig. 26 shows this matrix F_4 and its decomposition into the sum of sparse matrices $\alpha_0 * R0_4 + \alpha_1 * R1_4 + \alpha_2 * R2_4 + \alpha_3 * R3_4$ where $R0_4$, $R1_4$, $R2_4$ and $R3_4$ are unitary matrices from Fig. 6.

$$
\begin{array}{|c|c|c|c|}
\hline
\alpha_0 & -\alpha_1 & \alpha_2 & -\alpha_3 \\
\hline
\alpha_1 & \alpha_0 & -\alpha_3 & -\alpha_2 \\
\hline
\alpha_2 & -\alpha_3 & \alpha_0 & -\alpha_1 \\
\hline
-\alpha_3 & -\alpha_2 & \alpha_1 & \alpha_0 \\
\hline
\end{array}
=
$$

$$
= \alpha_0 \begin{vmatrix} 1&0&0&0 \\ 0&1&0&0 \\ 0&0&1&0 \\ 0&0&0&1 \end{vmatrix}
+ \alpha_1 \begin{vmatrix} 0&-1&0&0 \\ 1&0&0&0 \\ 0&0&0&-1 \\ 0&0&1&0 \end{vmatrix}
+ \alpha_2 \begin{vmatrix} 0&0&0&-1 \\ 0&0&-1&0 \\ 0&-1&0&0 \\ -1&0&0&0 \end{vmatrix}
+ \alpha_3 \begin{vmatrix} 0&0&1&0 \\ 0&0&0&-1 \\ 1&0&0&0 \\ 0&-1&0&0 \end{vmatrix}
$$

Fig. 26. The matrix F_4 of frequencies α_0, α_1, α_2, and α_3 of members 33, 32, 23, 22 of the H-doublets-alphabet and its decomposition into the sum $\alpha_0 R0_4 + \alpha_1 R1_4 + \alpha_2 R2_4 + \alpha_3 R3_4$, where $R0_4$, $R1_4$, $R2_4$ and $R3_4$ are unitary matrices from Fig. 6.

But the set of unitary matrices $R0_4$, $R1_4$, $R2_4$ and $R3_4$ is closed relative to multiplication in correspondence with the multiplication table of the algebra of split-quaternions by J. Cockle (Fig. 7). By this reason the frequency matrix F_4 represents a split-quaternion by J. Cockle having the following linear form of its representation by means of the expression (22):

$$F_4 = \alpha_0 * E + \alpha_1 * i + \alpha_2 * j + \alpha_3 * k \tag{22}$$

where E is the real unit and i, j, k – imaginary units of split-quaternions of J. Cockle.

Each long DNA sequence can be considered as a sequence of 4 H-doublets 33, 32, 23, 22 with certain values of their frequencies α_0, α_1, α_2, and α_3. It allows using the expression (22) of split-quaternions of J.Cockle to characterize and compare different long DNA sequences on the language of the algebra of these 4-dimensional hypercomplex numbers. This gives a new algebraic tool for comparison analysis of different DNA and RNA sequences in the field of evolutionary biology on the basis of properties of this algebra (https://en.wikipedia.org/wiki/Split-quaternion).

Let us turn now to the H-triplets-alphabet with its 8 members 333, 332, 323, 322, 233, 232, 223, 222. Frequencies of these members in a real DNA sequence will be denoted by symbols β_0, β_1, β_2, β_3, β_4, β_5, β_6, β_7 correspondingly. Fig. 22 shows locations of the members in the genetic (8*8)-matrix of 64 triplets. If in this matrix the all members 333, 332, 323, 322, 233, 232, 223, 222 are replaced by their frequencies β_0, β_1, β_2, β_3, β_4, β_5, β_6, β_7, the numeric matrix F_8 of the frequencies arises. Fig. 27 shows this matrix F_8 and its decomposition into the sum of sparse matrices $F_8 = \beta_0 * R8_0 + \beta_1 * R8_1 + \beta_2 * R8_2 + \beta_3 * R8_3 + \beta_4 * R8_4 + \beta_5 * R8_5 + \beta_6 * R8_6 + \beta_7 * R8_7$ where $R8_0$, $R8_1$, $R8_2$, $R8_3$, $R8_4$, $R8_5$, $R8_6$, $R8_7$ are unitary matrices from Fig. 17.

$$
\begin{array}{|c|c|c|c|c|c|c|c|}
\hline
\beta_0 & \beta_1 & -\beta_2 & -\beta_3 & \beta_4 & \beta_5 & -\beta_6 & -\beta_7 \\
\hline
\beta_1 & \beta_0 & -\beta_3 & -\beta_2 & \beta_5 & \beta_4 & -\beta_7 & -\beta_6 \\
\hline
\beta_2 & \beta_3 & \beta_0 & \beta_1 & -\beta_6 & -\beta_7 & -\beta_4 & -\beta_5 \\
\hline
\beta_3 & \beta_2 & \beta_1 & \beta_0 & -\beta_7 & -\beta_6 & -\beta_5 & -\beta_4 \\
\hline
\beta_4 & \beta_5 & -\beta_6 & -\beta_7 & \beta_0 & \beta_1 & -\beta_2 & -\beta_3 \\
\hline
\beta_5 & \beta_4 & -\beta_7 & -\beta_6 & \beta_1 & \beta_0 & -\beta_3 & -\beta_2 \\
\hline
-\beta_6 & -\beta_7 & -\beta_4 & -\beta_5 & \beta_2 & \beta_3 & \beta_0 & \beta_1 \\
\hline
-\beta_7 & -\beta_6 & -\beta_5 & -\beta_4 & \beta_3 & \beta_2 & \beta_1 & \beta_0 \\
\hline
\end{array}
=
$$

$$
\begin{vmatrix} 1&0&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0&0 \\ 0&0&1&0&0&0&0&0 \end{vmatrix}
\quad
\begin{vmatrix} 0&1&0&0&0&0&0&0 \\ 1&0&0&0&0&0&0&0 \\ 0&0&0&1&0&0&0&0 \end{vmatrix}
\quad
\begin{vmatrix} 0&0&-1&0&0&0&0&0 \\ 0&0&0&-1&0&0&0&0 \\ 1&0&0&0&0&0&0&0 \end{vmatrix}
\quad
\begin{vmatrix} 0&0&0&-1&0&0&0&0 \\ 0&0&-1&0&0&0&0&0 \\ 0&1&0&0&0&0&0&0 \end{vmatrix}
$$

$$= \beta_0 \begin{vmatrix} 0&0&0&1&0&0&0&0 \\ 0&0&0&0&1&0&0&0 \\ 0&0&0&0&0&1&0&0 \\ 0&0&0&0&0&0&1&0 \\ 0&0&0&0&0&0&0&1 \end{vmatrix} + \beta_1 \begin{vmatrix} 0&0&1&0&0&0&0&0 \\ 0&0&0&0&0&1&0&0 \\ 0&0&0&0&1&0&0&0 \\ 0&0&0&0&0&0&0&1 \\ 0&0&0&0&0&0&1&0 \end{vmatrix} + \beta_2 \begin{vmatrix} 0&1&0&0&0&0&0&0 \\ 0&0&0&0&0&0&-1&0 \\ 0&0&0&0&0&0&0&-1 \\ 0&0&0&0&1&0&0&0 \\ 0&0&0&0&0&1&0&0 \end{vmatrix} + \beta_3 \begin{vmatrix} 1&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&-1 \\ 0&0&0&0&0&0&-1&0 \\ 0&0&0&0&0&1&0&0 \\ 0&0&0&0&1&0&0&0 \end{vmatrix}$$

$$+ \beta_4 \begin{vmatrix} 0&0&0&0&1&0&0&0 \\ 0&0&0&0&0&1&0&0 \\ 0&0&0&0&0&0&-1&0 \\ 0&0&0&0&0&0&0&-1 \\ 1&0&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0&0 \\ 0&0&-1&0&0&0&0&0 \\ 0&0&0&-1&0&0&0&0 \end{vmatrix} + \beta_5 \begin{vmatrix} 0&0&0&0&0&1&0&0 \\ 0&0&0&0&1&0&0&0 \\ 0&0&0&0&0&0&0&-1 \\ 0&0&0&0&0&0&-1&0 \\ 0&1&0&0&0&0&0&0 \\ 1&0&0&0&0&0&0&0 \\ 0&0&0&-1&0&0&0&0 \\ 0&0&-1&0&0&0&0&0 \end{vmatrix} + \beta_6 \begin{vmatrix} 0&0&0&0&0&0&-1&0 \\ 0&0&0&0&0&0&0&-1 \\ 0&0&0&0&-1&0&0&0 \\ 0&0&0&0&0&-1&0&0 \\ 0&0&-1&0&0&0&0&0 \\ 0&0&0&-1&0&0&0&0 \\ -1&0&0&0&0&0&0&0 \\ 0&-1&0&0&0&0&0&0 \end{vmatrix} + \beta_7 \begin{vmatrix} 0&0&0&0&0&0&0&-1 \\ 0&0&0&0&0&0&-1&0 \\ 0&0&0&0&0&-1&0&0 \\ 0&0&0&0&-1&0&0&0 \\ 0&0&0&-1&0&0&0&0 \\ 0&0&-1&0&0&0&0&0 \\ 0&-1&0&0&0&0&0&0 \\ -1&0&0&0&0&0&0&0 \end{vmatrix}$$

Fig. 27. The matrix F_8 of frequencies $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7$ of 8 members 333, 332, 323, 322, 233, 232, 223, 222 of the H-triplets-alphabet and its decomposition into the sum $F_8 = \beta_0*R80+\beta_1*R81+\beta_2*R82+\beta_3*R83+\beta_4*R84+\beta_5*R85+\beta_6*R86+\beta_7*R87$ where $R_{80}, R_{81}, R_{82}, R_{83}, R_{84}, R_{85}, R_{86}, R_{87}$ are unitary matrices from Fig. 17.

But the set of 8 unitary matrices $R_{80}, R_{81}, R_{82}, R_{83}, R_{84}, R_{85}, R_{86}, R_{87}$ is closed relative to multiplication in correspondence with the multiplication table of the algebra of bi-split-quaternions (Fig. 18). By this reason the frequency matrix F_8 represents a bi-split-quaternion having the following linear form of its representation by means of the expression (23):

$$F_8 = \beta_0*E+\beta_1*i_1+\beta_2*i_2+\beta_3*i_3+\beta_4*i_4+\beta_5*i_5+\beta_6*i_6+\beta_7*i_7 \qquad (23)$$

where E is the real unit and $i_1, i_2, i_3, i_4, i_5, i_6$ – imaginary units of bi-split-quaternions.

Each long DNA sequence can be considered as a sequence of 8 H-triplets 333, 332, 323, 322, 233, 232, 223, 222 with certain values of their frequencies $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7$. It allows using the expression (23) of bi-split-quaternions to characterize and compare different long DNA sequences on the language of the algebra of these 8-dimensional hypercomplex numbers in addition to the above-mentioned algebra of 4-dimensional split-quaternions. The algebra of these 8-dimensional hypercomplex numbers can be extended further for cases of 2^n-dimensional hypercomplex numbers of a relevant type to use them for a comparative characterizing different long DNA sequences. One can suppose that such method of a comparative characterison and analysis by means of hypercomplex numbers can be used not only for long DNA sequences but also for genes and other relative short sequences of DNA and RNA.

But sets of members of the H-monoplets-alphabet, the H-doublets-alphabet, the H-triplets-alphabet, etc. can be characterized also in a connection with matrix representations of the algebra of hyperbolic numbers (or split-complex numbers, also double number) (https://en.wikipedia.org/wiki/Split-complex_number) and its 2^n-dimensional extensions. Fig. 28 shows the known matrix form of the repsesentation of these hyperbolic numbers d=a*1+b*j and its decomposition into 2 sparse matrices with the multiplication table for these matrices or basic units (here $j^2 = +1$; j is the imaginary unit; a, b are real coefficients). Both of these sparse matrices are unitary matrices.

$$d = a*E + b*j \;\blacktriangleright\; \begin{vmatrix} a, b \\ b, a \end{vmatrix} = a*\begin{vmatrix} 1, 0 \\ 0, 1 \end{vmatrix} + b*\begin{vmatrix} 0, 1 \\ 1, 0 \end{vmatrix} ;$$

*	1	j
1	1	j
j	j	1

Fig. 28. On the left: the matrix representation of hyperbolic numbers $d=a*1+b*j$ and its decomposition into two sparse matrices, which represent the real unit 1 (the first sparse matrix) and the imaginary unut j (the second sparse matrix). Here $j^2 = +1$, *a* and *b* are real numbers. On the right: the multiplication table of the basic elements of hyperbolic numbers.

The second tensor power of this matrix [a, b; b, a] gives the (4*4)-matrix [a, b; b, a]$^{(2)}$, which is the matrix representation of 4-dimensional hypercomplex numbers, the dyadic-shift decomposition of which gives 4 sparse unitary matrices with some real coefficient for each. The set of these sparse matrices is closed relative to multiplication and satisfies the multiplication table of the algebra of   4-dimensional hyperbolic matrions. The similar situation holds true for matrices [a, b; b, a]$^{(n)}$, n = 3, 4, 5 ,… . These questions are deacribed in detais in our books [Petoukhov, 2008; Petoukhov, He, 2009] about the algebra of 2-dimensional hyperbolic numbers and its extensions (by means of the tensor powers of their matrix representations) into algebras of 2^n-dimensional hypercomplex numbers called as hyperbolic matrions.

Taking this into account, let us return to the symbolic matrices in Fig. 2 and begin with the (2*2)-matrix [C, A; T, G]. If each of these nucleotides is represented by its number of hydrogen bonds (C=G=3 and A=T=2), then this matrix is transformed into the numeric matrix [3, 2; 2, 3]. If each of these numbers 3 and 2 of hydrogen bonds is replaced by its frequency q or p in the considered long DNA sequence, the frequency matrix [q, p; p, q] arises. In line with the expression in Fig. 28, this frequency matrix is the matrix representation of the hyperbolic number $q+p$*j, where j^2 = +1. This frequency (2*2)-matrix in its different tensor powers [q, p; p, q]$^{(n)}$ represents an appropriate 2^n-dimensional hyperbolic matrion. In the proposed model approach on the basis of hyperbolic matrions, entries of the matrix [q, p; p, q]$^{(n)}$ gives model sets of approximate frequencies for all members of a corresponding H-n-alphabet, members of which are components of the matrix [3, 2]$^{(n)}$. From this point of view, each long DNA sequence in relation of frequencies of its members of H-n-alphabets can be modeled on the basis of hyperbolic matrions for a comparison analysis with other DNA sequences.

The study of these questions is continued and its results will be published later.

Some concluding remarks

The article describes author's results about connections of genetic matrices with unitary matrices, the logical operation of the modulo-2 addition and complete orthogonal systems of functions. These genetic matrices represent genetic alphabets jointly with known features of the degeneracy of the genetic code. These results are interesting by the following main reasons.

Firstly, they give new approaches to model some genetic structures and phenomena on the basis of mathematical formalisms of quantum mechanics and quantum informatics where unitary operators have a key meaning. Correspondingly – from this standpoint - a hidden logic organization of the genetic system should be considered in the light of notions of quantum logic. Our results show that, from this modeling standpoint, the genetic system is a whole hierarchical system of interconnected unitary matrices of different orders woven together and formed tensor families of unitary matrices. Some of these unitary genetic matrices coincide with well-known quantum gates of quantum informatics; all other unitary genetic matrices can be also considered as special quantum gates for hidden quantum-information calculations in the genetic system. Complementary relations exist among some unitary genetic matrices.

We suppose that unitary genetic operators (unitary matrices) are the basis for calculations in genetics by some analogy with calculations in quantum informatics. In the frame of our model approach we put forward the working hypothesis that DNA- and RNA-sequences of n-plets (of doublets, triplets, etc.) serve to define unitary operators for quantum calculations in genetics by analogy with quantum-logical calculations in quatum computing. From this standpoint, DNA- and RNA-sequences are instruments to define systems of interconnected unitary operators for quantum calculations by means of the quantum logic (in particular, this is reflected in the special mosaic organisation of genetic matrices in Figs. 4 and 5). The presented materials about connections of genetic systems with quantum informatics (see additionally [Petoukhov, 2017a]) has led us to revealing hidden regularities (or rules) in long DNA sequences and complete sets of chromosomes of some organisms. These supposed and described rules concern frequencies (or probabilities) of members of H-n-alphabets of n-plets of DNA hydrogen bonds. The model of these phenomenological rules was proposed on the basis of the new notion of genetic qubits and formalisms of quantum informatics. Some analogies of these rules and of their model with the Hardy-Weinberg law and its equation were noted. Here one should note that the Hungarian scientist Gyorgy Darvas was the first who – in his study of quantum electrodynamics - paid

attention on connections of the genetic numeric matrices with Pauli's matrices [Darvas, Petoukhov, 2017]. It is additional interesting that cyclic shifts of positions in doublets and triplets transform the mosaic matrices in Figs. 4 and 5 into new mosaic matrices [Petoukhov, 2008; Petoukhov, He, 2009], which are connected with new systems of unitary genetic matrices.

Secondly, described unitary genetic matrices contain complete orthogonal systems of functions in their rows or columns. But it is known the following: "*after Fourier it was found that for some problems, harmonic sinusoids rather than other systems of orthogonal functions, for example, the Legendre polynomials, are better suited. In fact, any particular problem needs its own system of orthogonal functions. This was most clearly manifested in the course of the development of quantum mechanics*" [Soroko, 1973]. Correspondingly one can think that the genetic systems have their own orthogonal systems of functions, which should be used in physiology for appropriate spectral decompositions to study genetically inherited processes and structures (including genetic sequences, information processes in neuronal systems, cardio-vascular processes, etc.).

Thirdly, described fractal features of the mentioned tensor families of unitary genetic matrices give additional materials to the wide topic of inherited fractal-like structures in biological bodies, including symmetries in long texts of single stranded DNA [Petoukhov, 2017a] and facts about connections of fractals with cancer [Baish, Jain, 2000; Bizzarri et al, 2011; Dokukin et al, 2015; Lennon et al, 2015; Perez, 2017]. Fractal patterns are related with the theory of dynamic chaos, which has many applications in sciences and technology (see, for example, [Dmitriev, 2002; Potapov, 2015]). A specifity of fractal patterns in tensor families of unitary genetic matrices can be used for a further development of the theory of dynamic chaos and its applications. The bridge between knowledges about fractals in information technologies and in bio-information systems can lead to a mutual enrichment of both these fields.

Fourthly, we show the connection between the probabilities of members of alphabets of n-plets of hydrogen bonds in DNA sequences with some algebras of 2^n-dimensional hypercomplex numbers. This connection allows a comparison of different DNA sequences by means of their personal hypercompex numbers («hypercomplex passports» of DNA sequences) for studies in the field of evolutionary biology, medicine and biothechnology.

The author hopes that the further usage in genetics the concepts and formalisms of quantum informatics, which was undertaken in this article in the connection with unitary genetic matrices, will lead to the development of substantial quantum-information genetics. This will promote the inclusion of genetics and all biology in the field of profound mathematical natural science. Consideration of biological phenomena, including the phenomena of inheritance of the intellectual abilities of biological bodies, from the standpoint of the theory of quantum computers, gives many valuable opportunities for their comprehension and also for development of artificial intelligence systems [Petoukhov, 2016a,b; Petoukhov, Petukhova, 2017a; Petoukhov et al., 2017] (the work [Biamonte et al, 2017] contains a review about quantum computing and the problems of artificial intelligence). For example, an adult human organism has around 10 trillion (10^{14}) human cells and each of cells containts an identical complect of DNA, whose genetic information is used for physiological functioning organism as the holistic system of cells. How such huge number of cells can reliably functioning as a cooperative whole? Quantum informatics and associations with quantum computers can help to model and understand such holistic biological systems with their ability of computing complex tasks and transfering genetic information from one generation to another. Quantum-information approaches allow modeling complex biological systems without using data and hypotheses about interactions between adjacent molecules or between separate biological cells each with other; all of such separate elements are parts of a holistic organism as a quantum-information essence. The fundamental question about quantum computing was firstly touched upon in the book [Manin, 1980].

References

Annett M. (1985). *Left, right, hand and brain: the right shift theory*. New Jersey: Lawrence Erlbaum.

Ahmed N., Rao K. (1975). *Orthogonal transforms for digital signal processing.*   N-Y, Springer-Verlag Inc.

Annett M. (1992). Spatial ability in subgroupt of left- and right-handers. *British Journal of Psychology, 83*, 493-962.

Baish J.W., Jain R.K. (2000). Fractals and cancer. *Cancer Research*, 60, pp. 3683–3688, July 15 2000.

Biamonte J., Wittek P., Pancotti N., Rebentrost P., Wiebe N., Lloud S. (2017). Quantum machine learning. – *Nature,* v. 549, pages195–202 (14 September 2017), doi:10.1038/nature23474.

Bellman R. (1960). *Introduction to Matrix Analysis*. New-York: Mcgraw-Hill Book Company, Inc., 351 pp.

Bizzarri M., Giuliani A., Cucina A., Anselmi F. D., Soto A. M., Sonnenschein C. (2011). Fractal analysis in a Systems Biology approach to cancer. - *Semin Cancer Biol* . June, 21(3): pp. 175–182. doi:10.1016/j.semcancer.2011.04.002.

Chapeville F., Haenni A.-L. (1974). *Biosynthese des Proteines*. – Herman Collection, Paris Methodes.

Chargaff E. (1951). Structure and function of nucleic acids as cell constituents. - *Fed. Proc*., 10, 654-659.

Chargaff E. (1971). Preface to a Grammar of Biology: A hundred years of nucleic acid research. - *Science*, 172, 637-642

Chargaff E. (1975). A fever of reason. *Annu. Rev. Biochem*., 44:1-20.

Crow. J.f. (1999). Hardy, Weinberg and language impediments". *Genetics*, **152** (3): 821–825. ISSN 0016-6731. PMC 1460671 . PMID 10388804.

Darvas G., Petoukhov S.V. (2017). Algebra that demonstrates similitude between transformation matrices of genetic codes and quantum electrodynamics. - Presentation to the workshop organised 27.11.2017 at the «Competence center for algorithmic and mathematical methods in biology, biotechnology and medicine» of the Hochschule Mannheim, https://www.cammbio.hs-mannheim.de/fileadmin/user_upload/projekte/cammbio/events/20171127-kickoff/EP-Petoukhov.pdf

Derzhypolskyy A., Melenevskyy D., Gnatovskyy A. (2007). A Comparative Analysis of Associative Properties of Fourier versus Walsh Digital Holograms. - *Acta Physica Polonica A*, vol. 112, No. 5, pp. 1101-1106.

Dmitriev A.A. (2002). Design of message-carrying chaotic sequences // *Nonlinear phenomena in complex systems*, v. 5. № 1. pp. 78.

Dokukin M.E., Guz N.V., Woodworth C.D., Sokolov I. (2015). Emergence of fractal geometry on the surface of human cervical epithelial cells during progression towards cancer. - *New J Phys.,* Mar 10; 17(3). pii: 033019.

Dragovich B., Dragovich A. (2007). p-Adic Degeneracy of the Genetic Code. - https://arxiv.org/abs/0707.0764 .

Dragovich B., Khrennikov A., Misic N.Z. (2017). Ultrametrics in the genetic code and the genome. - *Applied Mathematics and Computation,* 309, 350–358.

Fimmel, E., Danielli A., Strüngmann L. (2013). On dichotomic classes and bijections of the genetic code. *J. Theor. Biol., 336*, 221–230.

Fimmel, E., Giannerini, S., Gonzalez, D., Strüngmann, L. (2014). Circular codes, symmetries and transformations. - *J. Math. Biol., 70*, 1623–1644.

Fimmel E., Strüngmann L. (2016). Yury Borisovich Rumer and his 'biological papers' on the genetic code. *Phil. Trans. R. Soc. A*, 374: 20150228.

Forsdyke D. R. (2006). Evolutionary Bioinformatics. New-York: Springer Verlag.

Frenkel I., Libine M. (2011). Split quaternionic analysis and separation of the series for SL(2,R) and SL(2,C)/SL(2,R). - *Advances in Mathematics*, v. 228, p. 678–763, https://ac.els-

cdn.com/S0001870811001897/1-s2.0-S0001870811001897-main.pdf?_tid=1aad0f3f-3dfe-4e0f-a5e3755ca0b56a7e&acdnat=1520069976_2e1a6bece7261f111bdce839410667ca].

Gazalé M. J. (1999). *Gnomon: from Pharaons to Fractals*. USA: Princeton Univ. Press, 280 p., ISBN-10: 0691005141

Gazzaniga M. S. (1995). Principles of human brain organization derived from split brain studies. *Neuron, 14*, 217-228.

Gillespie, J. H. (2004). *Population genetics : a concise guide* (2. ed.). Baltimore, Md.: The Johns Hopkins University Press. ISBN 0801880084

Harmuth, H. F. (1977). *Sequency theory. Foundations and applications*. N.-Y.: Academic Press.

Harmuth, H. F. (1989). *Information theory applied to space-time physics*. Washington: The Catholic University of America, DC.

Hellige J.B. (1993). *Hemispheric Asymmetry: What's Right and What's Left. Cambridge*. Massachusetts: Harvard University Press.

Hu Z.B., Petoukhov S.V. (2017). Generalized crystallography, the genetic system and biochemical esthetics. - *Structural Chemistry*, v. 28, №1, pp. 239-247. doi:10.1007/s11224-016-0880-0, http://link.springer.com/journal/11224/28/1/page/2

Hu Z.B., Petoukhov S.V., Petukhova E.S. (2017). I-Ching, dyadic groups of binary numbers and the geno-logic coding in living bodies. - *Progress in Biophysics and Molecular Biology* (in print, https://doi.org/10.1016/j.pbiomolbio.2017.08.018).

Igamberdiev A.U. (1993). Quantum mechanical properties of biosystems: a framework for complexity, structural stability, and transformations.- *Biosystems*, 31 (1), pp. 65–73.

Ivanov V.V. (1978). *Even and Odd: Asymmetry of the Brain and Sign Systems.* (*Chet i nechet: assimetria mozga i znakovyh system*). Moscow, Sovetskoie radio, 184 p. (in Russian).

Ji, S. (2015). Planckian distributions in molecular machines, living cells, and brains: The wave-particle duality in biomedical sciences, - *Proc. Intern. Conf. on Biology and Biomedical Engineering*. Vienna, March 15-17, pp. 115-137.

Ji S. (2017). *The Cell Language Theory: Connecting Mind and Matter*. - World Scientific Publishing Europe Ltd. , 450 p. ISBN-13: 978-1848166608.

Lennon F.E., Cianci G.C., Cipriani N.A., Hensing T.A., Zhang H.J., Chin-Tu Chen, Murgu S.D., Vokes E.E., Vannier M.W., Salgia R. (2015 November). Lung cancer - a fractal viewpoint. - *Nat Rev Clin Oncol*, 12(11):_pp. 664–675. doi:_10.1038/nrclinonc. 2015.108.

Manin Yu. I. (1980). *Computable and Uncomputable*, Moscow Sovetskoie Radio (in Russian).

Matsuno K. (1999). Cell motility as an entangled quantum coherence. – *BioSystems*, 51, pp. 15–19.

Matsuno K. (2003). Quantum mechanics in first, second and third person descriptions. – *BioSystems*, 68, pp. 107-118.

Matsuno K. (2015). On the physics of the emergence of sensorimotor control in the absence of the brain. - *Progress in Biophysics and Molecular Biology*, 119, pp. 313-323.

Matsuno K., Paton R.C. (2000). Is there a biology of quantum information? – *BioSystems*, 55, p.39–46

Moon Ho Lee, Han Haie, Sung Kook Lee (2017). A Mathematical Proof of Double Helix DNA to Reverse Transcription RNA for Bioinformatics. - Advances in Artificial Systems for Medicine and Education, Editors Zhengbing Hu, Sergey Petoukhov, Matthew He. Part of the Advances in Intelligent Systems and Computing book series (AISC, vol. 658), p. 53-67. https://link.springer.com/chapter/10.1007/978-3-319-67349-3_5

Morita Y., Sakurai Y. (1973). Holography by Walsh Waves. – Proceedings of the Symposium (4th) Held at the Catholic University of America, Washington, D. C. on 16-18 April, 1973, pp. 122-126.

Nielsen M.A., Chuang I.L. (2010). *Quantum Computation and Quantum Information*. New York: Cambridge University Press.

Okamura K., Wei J., Scherer S.W. (2007). Evolutionary implications of inversions that have caused intra-strand parity in DNA. - BMC Genomics, 8:160 doi:10.1186/1471-2164-8-160.

Patel A. (2001a). Quantum algorithms and the genetic code. Pramana – Journal of Physics, v. 56, 2-3, pp. 367-381, arXiv:quant-ph/0002037.

Patel A. (2001b). Testing quantum dynamics in genetic information processing. – Journal of Genetics, v. 80, 1, pp. 39-43.

Patel A. (2001c). Why genetic information processing could have a quantum basis. – Journal of Biosciences, v. 26, 2, pp. 145-151.

Pellionis A.J. (2008). The principle of recursive genome function. – *Cerebellum,* 7: pp. 348–359. DOI 10.1007/s12311-008-0035-y

Pellionisz A. J., Graham R., Pellionisz P. A. and Perez J. C. (2012). Recursive Genome Function of the Cerebellum: Geometric Unification of Neuroscience and Genomics, In: M. Manto, D.L. Gruol, J.D. Schmahmann, N. Koibuchi and F. Rossi, Eds., *Handbook of the Cerebellum and Cerebellar Disorders*, pp. 1381-1423.

Penrose R. (1989). The Emperor's New Mind. – Oxford University Press, USA.

Penrose R. (1996). *Shadows of the Mind: A Search for the Missing Science of Consciousness.* - Oxford University Press, USA, 480 p.

Perez J.C. (2010).   Codon populations in single-stranded whole human genome DNA are fractal and fine-tuned by the Golden Ratio 1.618. - *Interdiscip Sci Comput Life Sci* 2: pp. 228–240. DOI: 10.1007/s12539-010-0022-0

Perez J.-Cl. (2013). The "3 Genomic Numbers" Discovery: How Our Genome Single-Stranded DNA Sequence Is "Self-Designed" as a Numerical Whole. - *Applied Mathematics*, 4, 37-53, http://dx.doi.org/10.4236/am.2013.410A2004

Perez J.C. (2017). Sapiens Mitochondrial DNA Genome Circular Long Range Numerical Meta Structures are Highly Correlated with Cancers and Genetic Diseases mtDNA Mutations. - *J Cancer Sci Ther*, 9:6. DOI: 10.4172/1948-5956.1000469.

Petoukhov S.V. (2008) *Matrix genetics, algebras of the genetic code, noise immunity*. M., RCD, 316 p. (in Russian).

Petoukhov S.V. (2010a). The degeneracy of the genetic code and Hadamard matrices. - https://arxiv.org/abs/0802.3366 from 26 December 2010.

Petoukhov S.V. (2010b). Unitary symmetries and matrix operators of the genetic code. – *Proceedings of the VI Moscow International Conference on Operations Research*. MAKS Press, p. 159-160, http://io.cs.msu.su/ORM2010_abstracts.pdf

Petoukhov S.V. (2011). Matrix genetics and algebraic properties of the multi-level system of genetic alphabets. - *Neuroquantology,* 2011, Vol 9, No 4, p. 60-81, http://www.neuroquantology.com/index.php/journal/article/view/501

Petoukhov S.V. (2012) Symmetries of the genetic code, hypercomplex numbers and genetic matrices with internal complementarities. - *Symmetry: Culture and Science*, vol. 23, № 3-4, p. 275-301. http://symmetry.hu/scs_online/SCS_23_3-4.pdf

Petoukhov S.V. (2015a). Resonances and genetic biomechanics. - *Symmetry: Culture and Science*, т. 26, №3, p. 379-397,
 http://petoukhov.com/PETOUKHOV_IN_SCS_2015.pdf

Petoukhov S.V. (2015c). Resonances and genetic biomechanics. - *Symmetry: Culture and Science*, т. 26, №3, 2015, p. 379-397, http://petoukhov.com/PETOUKHOV_IN_SCS_2015.pdf

Petoukhov S.V. (2015d). The concept of resonances in genetics ("resonant bioinformatics"). - Selected Works of the VII International Congress "Weak and Superweak Fields and Radiation in Biology and Medicine", September 07-11, 2015, St. Petersburg, pp. 1-41, http://www.biophys.ru/lib/sci/resonance/460-resonance-00001 (in Russian).

Petoukhov S.V. (2016a). The system-resonance approach in modeling genetic structures. – *Biosystems*, January 2016, v. 139, p. 1-11, http://petoukhov.com/PETOUKHOV_ARTICLE_IN_BIOSYSTEMS.pdf .

Petoukhov S.V. (2016b). Symmetries of the genetic code, Walsh functions and the theory of genetic logical holography. - *Symmetry: Culture and Science*, v.27, №2, p. 95-98.

Petoukhov S.V. (2017a). The rules of long DNA-sequences and tetra-groups of oligonucleotides. – 148 p., https://arxiv.org/abs/1709.04943.

Petoukhov S.V. (2017b). I-Ching, dyadic groups of binary numbers and the geno-logic coding in living bodies. - *Progress in Biophysics and Molecular Biology*, v. 131, December 2017, pp. 354-368.

Petoukhov S.V., He M. (2009) *Symmetrical Analysis Techniques for Genetic Systems and Bioinformatics: Advanced Patterns and Applications*. Hershey, USA: IGI Global. 271 p.

Petoukhov S.V., Petukhova E.S. (2017a). Symmetries in genetic systems and the concept of geno-logical coding. - *Information*, 2017, *8*(1), 2; doi:10.3390/info8010002, http://www.mdpi.com/2078-2489/8/1/2/htm

Petoukhov S.V., Petukhova E.S. (2017b). Resonances and the quest for transdisciplinarity. - *Information Studies and the Quest for Transdisciplinarity*. Editors M. Burgin, W.Hofkirchner, World Scientific, p. p. 467-487.

Petoukhov S.V., Petukhova E.S., Hazina L.V., Svirin V.I., Stepanyan I.V., Silova T.V. (2017). Geno-logical coding, united-hypercomplex numbers and systems of artificial intelligence. – In the book "*Advances in Artificial Systems for Medicine and Education*", Editors Z.B. Hu, S.V.Petoukhov and M.He. Part of the "Advances in Intelligent Systems and Computing" book series (AISAC, volume 658), p. 2-13, Springer, https://link.springer.com/search?query=978-3-319-67348-6.

Petoukhov S.V., Svirin V.I. (2012). Fractal genetic nets and symmetry principles in long nucleotide sequences. - *Symmetry: Culture and Science*, vol. 23, № 3-4, p. 303-322. http://petoukhov.com/PETOUKHOV_SVIRIN_FGN.pdf .

Petoukhov S.V., Svirin V.I. (2018). The New Wide Class of Symmetries in Long DNA-Texts. Elements of Quantum-Information Genetics. - Biologia Serbica, v. 40, №1, p. 51, Special Edition, ISSN 2334 6590, UDK 57(051). Book of Abstracts, Belgrade Bioinformatics Conference 2018, 18-22 June, 2018, Belgrade, Serbia, page 51.

Potapov A.A. (2015). Chaos Theory, Fractals and Scaling in the Radar: A Look from 2015. - Chapter 12 in the book: *The Foundations of Chaos Revisited: From Poincaré to Recent Advancements* / Ed. Skiadas, C.- Switzerland, Basel: Springer Int. Publ., 2016. P. 195-218.- (ISBN 978-3-319-29701-9).

Prabhu, V. V. (1993). Symmetry observation in long nucleotide sequences. *Nucleic Acids Res.*, 21, 2797-2800.

Quantum Aspects of Life (2008). / editors Abbott D., Davies P.C.W., Pati A.K.; foreword by Sir Roger Penrose. ISBN-13: 978-1-84816-253-2.

Rapoport D. (2016a) Klein Bottle Logophysics, Self-reference, Heterarchies, Genomic Topologies, Harmonics and Evolution, Part I:Morphomechanics, Space and Time in Biology & Physics, Cognition, Non-Linearity and the Structure of Uncertainty, *Quantum Biosystems*, vol. 7, no.1, 2016, 1-72

Rapoport D. (2016b). Klein Bottle Logophysics, Self-reference, Heterarchies, Genomic Topologies and Evolution, Part II: Nonorientability, Cognition, Chemical Topology and Eversions in Nature, *Quantum Biosystems*, vol. 7, no.1, 2016, 74-106.

Rapoport D. (2016c). Klein Bottle Logophysics, Self-reference, Heterarchies, Genomic Topologies and Evolution. Part III: The Klein Bottle logic of Genomics and its Dynamics, Quantum Information, Complexity and Palindromic Repeats in Evolution. *Quantum Biosystems,* vol. 7, no. 1, 107-172.

Rapoport A.E., Trifonov E.N. (2012): Compensatory nature of Chargaff's second parity rule. *Journal of Biomolecular Structure and Dynamics*, November, 1-13, DOI:10.1080/07391102.2012.736757

Rumer Yu. B. (1968). Systematization of the codons of the genetic code. - *Doklady Akademii Nauk SSSR,* vol. 183(1), p. 225-226 (in Russian).

Rumer Yu.B., Fet A.I. (1970). The Theory of Unitary Symmetry (Teoria unitarnoi simmetrii). – Moscow, Nauka, 400 pages (in Russian).

Seberry J., Wysocki D.J., Wysocki T.A. (2005). Some applications of Hadamard matrices. *Metrika,* 62, p. 221-239. https://www.uow.edu.au/~jennie/WEBPDF/2005_12.pdf

Simeonov P.L. (2013). On Some Recent Insights in Integral Biomathics. - *J. Progress in Biophysics and Molecular Biology*. Special Theme Issue on Integral Biomathics: Can Biology Create a Profoundly New Mathematics and Computation? Elsevier. ISSN: 0079-6107. DOI: 10.1016/j.pbiomolbio.2013.06.001. http://arxiv.org/abs/1306.2843.

Soroko L.M. (1973). Walsh's functions in physics and technics. - *Successes of Physical Sciences (Uspehi Fizicheskih Nauk)*, v. 111, issue 3.

Soroko L.M. (1974). "Experience" of Young-Walsh and Logical Holography. - *Materials of 6th national conference on holography* (11-17.02.1974, Yerevan), Leningrad Institute of Nuclear Physics, http://bsfp.media-security.ru/school6/5.htm.

Stambuk N. (1999). Circular coding properties of gene and protein sequences. *Croat. Chem. Acta, 72,* 999–1008.