Symmetry: Culture and Science
Vol. 30, No. 2, 161-179, 2019
https://doi.org/10.26830/symmetry_2019_2_161

SYMMETRY: SCIENCE AND ART

SYMMETRIES OF DNA ALPHABETS AND QUANTUM INFORMATIONAL FORMALISMS

Sergey V. Petoukhov[1*], Elena S. Petukhova[1], Vitaly I. Svirin[1]

[1] Mechanical Engineering Research Institute, Russian Academy of Sciences, M.Kharitonievsky per., 4, Moscow, 101990, Russia
E-mail: spetoukhov@gmail.com, http://petoukhov.com/.
[*]corresponding author

Abstract: *One of creators of quantum mechanics P. Jordan in his work on quantum biology claimed that life's missing laws were the rules of chance and probability of the quantum world. Correspondingly the article is devoted to some phenomenological symmetries in probabilities of components in long DNA sequences and to opportunities of their modelling by means of formalisms of quantum informatics. Using binary-oppositional traits of DNA nitrogenous bases, a notion "genetic qubit" is proposed and used for such modelling. Presented results support thoughts of many authors about quantum computing in biological organisms.*

Keywords: symmetry, DNA, binary opposition, genetic qubit, quantum informatics.

1 INTRODUCTION

A great discovery of the twentieth century physics was the probabilistic nature of physical phenomena at atomic scales, described by quantum mechanics. Informational molecules of heredity DNA and RNA belong to the microworld of quantum mechanics and they should obey the principles of quantum mechanics. Many authors have supposed that living organisms use principles of quantum informatics (Abbott, Davies, Pati, 2008; Altaisky, Filatov, 2001; Fimmel et al, 2019; Hu Z.B., Petoukhov, Petukhova,

2017a,b, 2018; Igamberdiev, 1993, 2004, 2007, 2008; Igamberdiev, Shklovskiy-Kordi, 2016, 2017; Josephson, 2018; Matsuno, 1999, 2003; Matsuno, Paton, 2000; Mikheenko, 2018; Patel, 2001a,b,c; Penrose,1996, 2019; Petoukhov, 2016, 2018a,b, 2019a,b). Genetic informatics is written in DNA molecules by means of sequences of four nitrogenous bases, which are frequently called letters of the DNA alphabet of monoplets (or of nucleotides): adenine A, cytosine C, guanine G and thymine T (RNA uses uracil U instead of thymine). Concatenations of these 4 letters form the alphabet of 16 doublets (AA, AC,…), the alphabet of 64 triplets (AAA, ACG,…), etc, which participate in the genetic coding systems. One of the creators of quantum mechanics P. Jordan in his work on quantum biology claimed that life's missing laws were the rules of chance and probability of the quantum world (Jordan, 1932; McFadden, Al-Khalili, 2018). From the standpoint of Jordan's statement, the study of probabilities or frequencies of nucleotide n-plets (monoplets, doublets, triplets, etc.) in long DNA sequences is important for discovering hidden biological laws and for developing quantum genetics. Recent articles (Darvas, 2018; Fimmel et al., 2019; Petoukhov, 2018b; Petoukhov, Petukhova, Svirin, 2019) were devoted to evidences in favour of existence of rules of new symmetries in probabilities of nucleotide *n*-plets (n = 1, 2, 3, 4, …) of long DNA sequences in eukaryotic and prokaryotic genomes.

This article shows additional opportunities to introduce formalisms of quantum informatics on the basis of symmetric peculiarities of DNA alphabets for modeling some genetic phenomena including the mentioned rules of symmetries of probabilities of nucleotide *n*-plets in genomes.

2 BINARY OPPOSITIONS IN THE DNA ALPHABET

Science does not know why the nucleotide alphabet of DNA has been created by nature from just four letters (A, T, C, 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 by means of their symmetrical peculiarities into the united molecular ensemble with its three pairs of binary-oppositional traits or indicators (Fig. 1):

(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, as above, 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.

№	Binary Symbols	C	A	G	T/U
1	$\mathbf{0}_1$ — pyrimidines $\mathbf{1}_1$ — purines	$\mathbf{0}_1$	$\mathbf{1}_1$	$\mathbf{1}_1$	$\mathbf{0}_1$
2	$\mathbf{0}_2$ — amino $\mathbf{1}_2$ — keto	$\mathbf{0}_2$	$\mathbf{0}_2$	$\mathbf{1}_2$	$\mathbf{1}_2$
3	$\mathbf{0}_3$ — three hydrogen bonds $\mathbf{1}_3$ — two hydrogen bonds	$\mathbf{0}_3$	$\mathbf{1}_3$	$\mathbf{0}_3$	$\mathbf{1}_3$

Figure 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.

Accordingly, each of 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 logic operation of modulo-2 addition denoted by the symbol $\oplus$: modulo-2 addition of any two such binary representations of the DNA-sequence coincides with the third binary representation of the same DNA-sequence: for example, $101011110 \oplus 100110011 = 001101101$. One can here remind 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 quantum informatics (Nielsen, Chuang, 2010)).

Taking into account the phenomenological fact that each of the 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 as corresponding pairs of binary symbols, for example, from the standpoint of two first binary-oppositional indicators. It is convenient for us for further description to show at the first position of each of the 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 the

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 a 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 the 64 triplets is represented as 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 *n*-plet is represented as a 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), whose 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-purine-purine” and so this column is correspondingly numerated 011). By contrast, entries of each row 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. 1), 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

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

Here one can remind a historical fact that the same principle of constructing square tables with quite similar binary numerations of their columns and rows was used else in the Ancient Chinese book «I-Ching», which was written a few thousand years ago and which is the most ancient historical example of systematic usage of binary numbers. The famous table of 64 hexagrams in Fu-Xi's order exists there, which has many deep analogies with the genetic matrix of 64 triplets (Hu, Petoukhov, Petukhova, 2017b; Petoukhov, 1999, 2008, 2017; Petoukhov, He, 2009). The ancient Chinese claimed that this table is the universal archetype of nature. They knew nothing about the genetic code, but the genetic code is constructed in line with the "I-Ching".

One should emphasize that these 3 separate genetic tables (Fig. 2) form the joint tensor (!) family of matrices since they are interrelated by the known operation of the tensor (or Kronecker) product of matrices. By definition, under tensor multiplication of two matrices, each entry of the first matrix multiplies 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 matrix of 16 doublets; the third tensor power of the matrix of the same matrix of 4 DNA-letters gives the matrix of 64 triplets with the same strict arrangement of entries (Fig. 3).

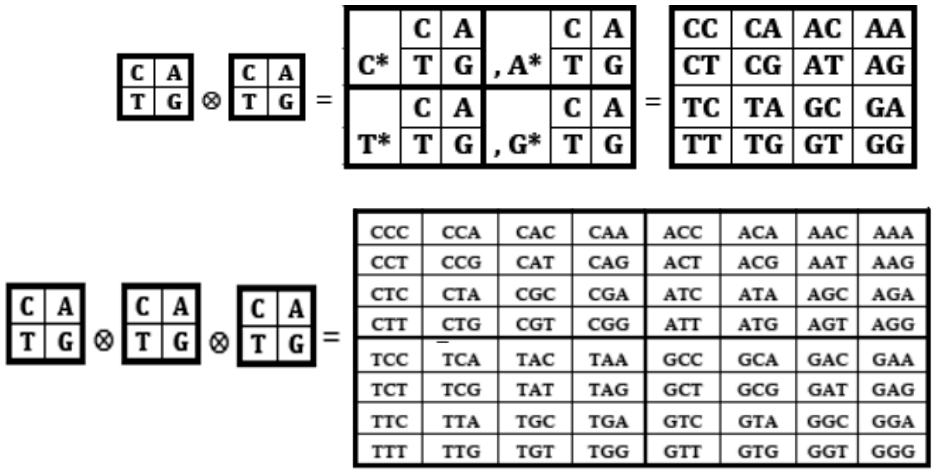

Figure 3: The tensor family of genetic matrices $[C, A; T, G]^{(n)}$ (here tensor power n = 1, 2, 3) of DNA-alphabets of 4 nucleotides, 16 doublets and 64 triplets. The symbol ⊗ means the tensor product.

One can see from a comparison of Fig. 3 with Fig. 2 that the tensor powers of the genetic matrix of 4 letters [C, A; T, G] produce - in a new way - the same square tables of the DNA-alphabets with the same strict arrangements of alphabetic components without any mention about binary-oppositional traits of genetic letters C, A, T and G, which were used to construct the square tables in Fig. 2.

So, the structural organization of the system of DNA-alphabets is connected with the algebraic operation of the tensor product. 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 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 basis instruments in quantum informatics.

3 THE DNA ALPHABET AND GENETIC QUBITS

This Section describes a possibility of using features of DNA alphabets in the construction of 2*n*-qubit systems in so called separable pure states for model tasks of mathematical genetics. Typical notions, denotations and formalisms of quantum informatics from the fundamental book (Nielsen, Chuang, 2010) are used. In particular we use the notion of quantum bits (or qubits) for model representations of elements of

single stranded DNA. We also use ordinary Dirac notations for vectors and operations with them:
$|\psi\rangle$ means a vector, which is known also as a ket-vector;
$\langle\psi|$ means a vector, which is dual to $|\psi\rangle$ and known as a bra-vector;
$\langle\varphi|\psi\rangle$ means scalar (or inner) product between the vectors $\langle\varphi|$ and $|\psi\rangle$;
$|\varphi\rangle \otimes |\psi\rangle$ means the tensor product of $|\varphi\rangle$ and $|\psi\rangle$;
$|\varphi\rangle|\psi\rangle$ is abbreviated notation for tensor product of $|\varphi\rangle$ and $|\psi\rangle$;
M^T is the transpose of the M matrix;
$\langle\varphi|M|\psi\rangle$: scalar product between $|\varphi\rangle$ and $M|\psi\rangle$.

In quantum informatics, such vector spaces H are considered, which are equipped with the scalar (or inner) product (so called Hilbert spaces). Let H_1 and H_2 be quantum mechanical state spaces, that is, finite dimensional Hilbert spaces with orthonormal basis states $|\alpha_i\rangle$ and $|\beta_j\rangle$, where $i = 1, \ldots, n$ and $j = 1, \ldots, m$. By a postulate of quantum mechanics, the state space of the composite system is given by the tensor product $H_1 \otimes H_2$ with basis states $\{|\alpha_i\rangle \otimes |\beta_j\rangle\}$, or in more compact notation $\{|\alpha_i\beta_j\rangle\}$. "*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* $\rho_1 \otimes \rho_2 \otimes \ldots \rho_n$" [Nielsen, Chuang, 2010, p. 102] (here ρ is density operator). If a quantum state can be represented as a vector of a Hilbert space, such state is called a pure quantum state. If a pure state $|\psi\rangle \subset H_1 \otimes H_2$ can be written in the form $|\psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle$, where $|\psi_i\rangle$ is a pure state of the i-th subsystem, it is said to be separable. Otherwise it is called entangled.

As known, a quantum bit (or qubit) is a unit of quantum information. For two-level quantum systems used as qubits, the state $|0\rangle$ is identified with the vector (1, 0), and similarly the state $|1\rangle$ with the vector (0, 1). Two possible states for a qubit are the states $|0\rangle$ and $|1\rangle$, 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 $|0\rangle$ or $|1\rangle$. It is possible to form linear combinations of states, often called superpositions (1):

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle, \quad (1)$$

The symbol $|\psi\rangle$ means a state of a qubit. The numbers α and β 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 standard notation for states in quantum mechanics is the Dirac notation "| >". The special states $|0\rangle$ and $|1\rangle$ 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 α and β. 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 $|\alpha|^2$, or the result 1, with probability $|\beta|^2$. Naturally, $|\alpha|^2 + |\beta|^2 = 1$, since the probabilities must sum to one. Geometrically, we can interpret this as the condition that the qubit's state be normalized to length 1. Values α and β are called amplitudes of probabilities. 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 the 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 $|x_1x_2\ldots x_n>$; a quantum state of such a system is specified by 2^n amplitudes [Nielsen, Chuang, 2010, p. 17]. In our model approach we interpret DNA-texts as quantum systems of many 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 of 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 qubits on the basis of different pairs of binary-oppositional indicators of adenine A, guanine G, cytosine C, and thymine T, which were shown above in Fig. 1. As we noted, each of these DNA bases can be uniquely defined by any two kinds of mentioned binary-oppositional indicators (Fig. 1). By analogy with the previous Section, to characterize each of the DNA-letters C, A, T, G we will use the first kind of indicators ("pyrimidine or purine") and the second kind of indicators ("three hydrogen bonds or two hydrogen bonds"). On the basis of each of these pairs of binary-oppositional indicators, a corresponding two-level quantum system can be formally introduced with a definition of its appropriate qubit (those qubits, which are introduced in genetic informatics on the basis of binary-oppositional indicators of genetic molecules, can be conditionally called "genetic qubits" or briefly "g-qubits").

Let us introduce, firstly, the notion of a genetic qubit as a two-level quantum system on the basis of the indicators "pyrimidine or purine": in this quantum system one level corresponds to the indicator "pyrimidine" and the second level – to the oppositional

indicator “purine”. 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_1. One can assume that the state $|0\rangle$ corresponds to the state “pyrimidine”, and the state $|1\rangle$ - to the state “purine”. By analogy with the expression (6), a state of such genetic qubit can be expressed by the expression (2), where α_0 and α_1 are amplitudes of probabilities of these computational basis states “pyrimidine” and “purine”:

$$|\psi_1\rangle = \alpha_0 |0\rangle + \alpha_1 |1\rangle, \qquad \alpha_0^2 + \alpha_1^2 = 1 \tag{2}$$

In a particular case, a qubit (2) can represent a pure state of a quantum system in a form of a sequence, which consists of pyrimidines and purines.

Secondly, let us introduce the notion of another genetic qubit as a two-level quantum system on the basis of the indicators “three hydrogen bonds or two hydrogen bonds”: in this quantum system one level corresponds to the indicator “three hydrogen bonds” and the second level – to the indicator “two hydrogen bonds”. In other words, such genetic qubit is represented by these two indicators and the state of such qubit is a vector in its appropriate 2-dimensional Hilbert space H_2. One can assume that the state $|0\rangle$ corresponds to the state “three hydrogen bonds”, and the state $|1\rangle$ - to the state “two hydrogen bonds”. By analogy with the expression (1), a state of such genetic qubit can be expressed by the expression (3), where β_0 and β_1 are amplitudes of probabilities of these computational basis states:

$$|\psi_2\rangle = \beta_0 |0\rangle + \beta_1 |1\rangle, \qquad \beta_0^2 + \beta_1^2 = 1 \tag{3}$$

In a particular case, the qubit (3) can represent a pure state of a quantum system in a form of a sequence, which consists of elements with three and two hydrogen bonds.

So we have two different 2-dimensional Hilbert spaces H_1 and H_2, to which pure states of genetic qubits (2) and (3) belong correspondingly. In our genetic case, the tensor product of the two-dimensional Hilbert space $H_1 \otimes H_2$ gives one four-dimensional Hilbert space with the following separable pure state of a quantum 2-qubit system:

$$|\psi_{12}\rangle = |\psi_1\rangle \otimes |\psi_2\rangle = (\alpha_0 |0\rangle + \alpha_1 |1\rangle) \otimes (\beta_0 |0\rangle + \beta_1 |1\rangle) =$$

$$\alpha_0\beta_0 |00\rangle + \alpha_0\beta_1 |01\rangle + \alpha_1\beta_0 |10\rangle + \alpha_1\beta_1 |11\rangle \tag{4}$$

Such 2-qubit system has four computational basis states denoted |00>, |01>, |10>, |11>. The amplitudes $\alpha_0\beta_0$, $\alpha_0\beta_1$, $\alpha_1\beta_0$ and $\alpha_1\beta_1$ of probabilities satisfy the normalization condition: $(\alpha_0\beta_0)^2 + (\alpha_0\beta_1)^2 + (\alpha_1\beta_0)^2 + (\alpha_1\beta_1)^2 = 1$. One can note that two kinds of indicators "pyrimidine-purine" and "three hydrogen bonds or two hydrogen bonds" for separate nitrogenous bases C, T, G, A define the following correspondence:

- Cytosine C corresponds to the computational basis state |00> of the 2-qubit system (4) since cytosine C is characterized by the indicator "pyrimidine", which is the computational basis state |0> in the first qubit (7), and also by the indicator "three hydrogen bonds", which is the computational basis state |0> in the second qubit (3). In the four-dimensional Hilbert space $H_1 \otimes H_2$, these computational basis states |0> and |0> of two genetic qubits (2) and (3) define the computational basis state |00> of the 2-qubit system (4);

- Thymine T corresponds to the computational basis state |01> of the 2-qubit system (4) since thymine T is characterized by the indicator "pyrimidine", which is the computational basis state |0> in the first qubit (5), and also by the indicator "two hydrogen bonds", which is the computational basis state |1> in the second qubit (3). In the four-dimensional Hilbert space $H_1 \otimes H_2$, these computational basis states |0> and |1> of two genetic qubits (5) and (3) define the computational basis state |01> of the 2-qubit system (4);

- Guanine G corresponds to the computational basis state |10> of the 2-qubit system (4) since guanine G is characterized by the indicator "purine", which is the computational basis state |1> in the first qubit (2), and also by the indicator "three hydrogen bonds", which is the computational basis state |0> in the second qubit (3). In the four-dimensional Hilbert space $H_1 \otimes H_2$, these computational basis states |1> and |0> of two genetic qubits (2) and (3) define the computational basis state |10> of the 2-qubit system (4).

- Adenine A corresponds to the computational basis state |11> of the 2-qubit system (4) since adenine A is characterized by the indicator "purine", which is the computational basis state |1> in the first qubit (2), and also by the indicator "two hydrogen bonds", which is the computational basis state |1> in the second qubit (3). In the four-dimensional Hilbert space $H_1 \otimes H_2$, these computational basis states |1> and |1> of two genetic qubits (2) and (3) define the computational basis state |11> of the 2-qubit system (4).

By this way, members of the tetra-group of nucleotides C, T, G and A get their representations as computational basis states of a 2-qubit system (9) in a four-dimensional Hilbert space $H_1 \otimes H_2$ with conditional denotations $|C\rangle=|00\rangle$, $|T\rangle=|01\rangle$, $|G\rangle=|10\rangle$ and $|A\rangle=|11\rangle$ (these computational basis states can be connected with pairs of photons of different frequencies, radiated by molecular elements of these dual indicators (Petoukhov, 2018b)). Therefore the expression (4) can be rewritten in the following conditional form (5):

$$|\psi_{12}\rangle = \alpha_0\beta_0\,|00\rangle + \alpha_0\beta_1\,|01\rangle + \alpha_1\beta_0\,|10\rangle + \alpha_1\beta_1\,|11\rangle =$$

$$= \alpha_0\beta_0\,|C\rangle + \alpha_0\beta_1\,|T\rangle + \alpha_1\beta_0\,|G\rangle + \alpha_1\beta_1\,|A\rangle \qquad (5)$$

We call the 2-qubit quantum system with its separable pure state (5) as a "monoplet CTGA-system". In a particular case, the state (5) of a 2-qubits monoplet CTGA-system can represent a separable pure state of a quantum system, which is a sequence of elements, where each element is pyrimidine or purine and, simultaneously, it has three or two hydrogen bonds by analogy with nucleotides. From (5) the probabilities $P(C)$, $P(T)$, $P(G)$ and $P(A)$ of computational basis states of separate nucleotides C, T, G and A are equal to the following:

$$P(C) = (\alpha_0\beta_0)^2,\ P(T) = (\alpha_0\beta_1)^2,\ P(G) = (\alpha_1\beta_0)^2,\ P(A) = (\alpha_1\beta_1)^2 \qquad (6)$$

They should satisfy the normalization condition (7):

$$(\alpha_0\beta_0)^2 + (\alpha_0\beta_1)^2 + (\alpha_1\beta_0)^2 + (\alpha_1\beta_1)^2 = 1 \qquad (7)$$

4 USING GENETIC QUBITS IN MODELING SOME SYMMETRIES IN LONG DNA SEQUENCES

This Section shows an example of using the introduced notion of genetic qubits to model some phenomenological symmetries in long DNA sequences.

DNA-texts of different organisms are represented in GenBank that contains hundreds of millions of sequences for more than one hundred thousand organisms. Modern technologies allow determining sequences of letters in DNA molecules of different organisms but there arises a problem: "What will we have when these genomic sequences are determined? … We are in the position of Johann Kepler when he first

began looking for patterns in the volumes of data that Tycho Brahe had spent his life accumulating " (Fickett & Burks, 1989).

Until recently, science acknowledged the second parity rule of E. Chargaff that stated approximate equality of probabilities for DNA-letters A and T, as well as also C and G in long nucleotide sequences of single stranded DNA. This is known in the literature as a symmetry principle and is connected with "a grammar of biology" [a term from (Chargaff, 1971)]. Many works of different authors are devoted to this second Chargaff's rule [see references in (Petoukhov, 2018b)]. Chargaff and his followers concentrated their attention on the comparison of individual probabilities of separate *n*-plets in DNA.

By contrast to this, the work (Petoukhov, 2018b) studied not individual probabilities of separate *n*-plets but collective (or total) probabilities of alphabetical clusters (or unions) of *n*-plets (that is doublets, triplets, etc.) in long DNA-texts. Just in this way, an unknown class of hidden symmetries and their rules for long DNA-texts was discovered and formulated. It sounds in short: a complete alphabet of *n*-plets with fixed length "*n*" contains 4^n members. Such alphabet can be considered as a union of 4 equal sub-alphabets, each of which contains 4^{n-1} members with one of the 4 letters A, T, C, G at a certain position (each of the n positions is connected with its own union of 4 sub-alphabets). By definition, in a DNA-text, the collective probability P_n (X_k) of all *n*-plets in each of such 4 sub-alphabets is equal to the sum of individual probabilities of its *n*-plets (here X means A, T, C or G; n = 1, 2, 3, 4, 5,.. is not too large; index k denotes the position inside *n*-plets, $k \leq n$).

Let us turn to the following rule formulated in (Petoukhov, 2018b) about a generalized symmetry for collective probabilities of sub-alphabets of *n*-plets in long DNA-texts:

- In long sequences of *n*-plets of single stranded DNA, the collective probabilites $P_n(X_k)$ of the sub-alphabets of *n*-plets, which have the letter X in their position k, are approximately equal to the individual probability $P(X)$ of the nucleotide X, regardless of the values of n and k.

Now we show how the introduced notion of genetic qubits allows modelling this phenomenological rule. We begin with the simplest case of long sequences of doublets of single stranded DNA (cases of long DNA sequences of triplets, tetraplets, etc. are modelled by analogy with this case).

To consider the case of doublets CC, CT, CG, CA, TC, TT, TG, TA, GC, GT, GG, GA, AC, AT, AG, AA one should accordingly expand the 4-dimensional Hilbert space $H_1 \otimes H_2$, considered in the previous Section, to the 16-dimensional Hilbert space $H_1 \otimes H_2 \otimes H_3 \otimes H_4$. Here the spaces H_1 and H_2 are related with the first letters of doublets and the spaces H_3 and H_4 are related with the second letters of doublets. These spaces H_3 and H_4 are defined in a close analogy with the spaces H_1 and H_2. Let us explain this.

By analogy with the expressions (1, 2), each of four letters C, T, G, A at the second position of doublets is interpreted firstly as a two-level quantum system on the basis of the oppositional indicators “pyrimidine or purine”: in this quantum system one level corresponds to the indicator “pyrimidine” and the second level – to the oppositional indicator “purine”. In other words, a new genetic qubit arises on the basis of these oppositional indicators for the second letters inside 16 doublets; the state of such qubit is a vector in its appropriate two-dimensional Hilbert space H_3. One can assume that the state $|0\rangle$ corresponds to the state «pyrimidine», and the state $|1\rangle$ - to the state “purine”. By analogy with the expression (2), a state of such genetic qubit can be denoted by the expression (8), where α_2 and α_3 are amplitudes of probabilities of these computational basis states “pyrimidine” and “purine” for the second letters in 16 doublets:

$$|\psi_3\rangle = \alpha_2 |0\rangle + \alpha_3 |1\rangle, \qquad \alpha_2^2 + \alpha_3^2 = 1 \tag{8}$$

Secondly, for the second letters of 16 doublets, the notion of another genetic qubit - as a two-level quantum system on the basis of the oppositional indicators “three hydrogen bonds or two hydrogen bonds” - is defined by analogy with the expression (3). In this quantum system, one level corresponds to the indicator “three hydrogen bonds” and the second level – to the indicator “two hydrogen bonds”. In other words, such genetic qubit is represented by these two indicators and the state of such qubit is a vector in its appropriate 2-dimensional Hilbert space H_4. One can assume that the state $|0\rangle$ corresponds to the state “hree hydrogen bonds”, and the state $|1\rangle$ - to the state “two hydrogen bonds”. By analogy with the expression (3), a state of such genetic qubit can be expressed by the expression (9), where β_2 and β_3 are amplitudes of probabilities of these computational basis states:

$$|\psi_4\rangle = \beta_2 |0\rangle + \beta_3 |1\rangle, \qquad \beta_2^2 + \beta_3^2 = 1 \tag{9}$$

In such a way, in the case of long DNA-texts of doublets, we have two additional 2-dimensional Hilbert spaces H_3 and H_4, to which pure states of genetic qubits (8) and (9) belong correspondingly. The tensor product of the two-dimensional Hilbert space

$H_3 \otimes H_4$ gives one four-dimensional Hilbert space with the following separable pure state of a quantum 2-qubit system for letters at the second position inside 16 doublets:

$$|\psi_{34}> = |\psi_3> \otimes |\psi_4> = (\alpha_2 |0> + \alpha_3 |1>) \otimes (\beta_2 |0> + \beta_3 |1>) =$$

$$\alpha_2\beta_2 |00> + \alpha_2\beta_3 |01> + \alpha_3\beta_2 |10> + \alpha_3\beta_3 |11> =$$

$$\alpha_2\beta_2|C> + \alpha_2\beta_3 |T> + \alpha_3\beta_2 |G> + \alpha_3\beta_3 |A> \qquad (10)$$

Such new 2-qubit system has four computational basis states denoted |00>, |01>, |10>, |11>. By analogy with the case of the Hilbert space $H_1 \otimes H_2$ for the first letters of doublets with conditional denotations |C>=|00>, |T>=|01>, |G>=|10> and |A>=|11> (see the expression (5)), in the Hilbert space $H_3 \otimes H_4$ similar denotations are used for 2-qubit systems of the second letters of doublets: |C>=|00>, |T>=|01>, |G>=|10> and |A>=|11>. The amplitudes $\alpha_2\beta_2$, $\alpha_2\beta_3$, $\alpha_3\beta_2$ and $\alpha_3\beta_3$ of probabilities in (10) satisfy the normalization condition (11):

$$(\alpha_2\beta_2)^2 + (\alpha_2\beta_3)^2 + (\alpha_3\beta_2)^2 + (\alpha_3\beta_3)^2 = 1. \qquad (11)$$

In the 16-dimensional Hilbert space $H_1 \otimes H_2 \otimes H_3 \otimes H_4$ for the case of 16 doublets, we have in our model approach the following 16 computational basis states with their appropriate amplitudes of probabilities in a long DNA-text:

$$|\psi_{12}> \otimes |\psi_{34}> = \alpha_0\beta_0\alpha_2\beta_2|CC> + \alpha_0\beta_0\alpha_2\beta_3|CT> + \alpha_0\beta_0\alpha_3\beta_2|CG> + \alpha_0\beta_0\alpha_3\beta_3|CA> +$$
$$\alpha_0\beta_1\alpha_2\beta_2|TC> + \alpha_0\beta_1\alpha_2\beta_3|TT> + \alpha_0\beta_1\alpha_3\beta_2|TG> + \alpha_0\beta_1\alpha_3\beta_3|TA> +$$
$$\alpha_1\beta_0\alpha_2\beta_2|GC> + \alpha_1\beta_0\alpha_2\beta_3|GT> + \alpha_1\beta_0\alpha_3\beta_2|GG> + \alpha_1\beta_0\alpha_3\beta_3|GA> +$$
$$\alpha_1\beta_1\alpha_2\beta_2|AC> + \alpha_1\beta_1\alpha_2\beta_3|AT> + \alpha_1\beta_1\alpha_3\beta_2|AG> + \alpha_1\beta_1\alpha_3\beta_3|AA> \qquad (12)$$

In the state (12) of a 4-qubit "doublet CTGA-system", 16 computational basis states are represented by 16 doublets: |0000>=|CC>, |0001>=|CT>, |0010>=|CG>, |0011>=|CA>, |0100>=|TC>, |0101>=|TT>, |0110>=|TG>, |0111>=|TA>, |1000>=|GC>, |1001>=|GT>, |1010>=|GG>, |1011>=|GA>, |1100>=|AC>, |1101>=|AT>, |1110>=|AG>, |1111>=|AA>. In our model approach, these 16 computational basis states are interpreted as representations of appropriate 16 genetic doublets.

From the expressions (12) and (6), we have the following collective probability $P_2(C_1)$ of all 4 doublets having the first letter C in them (they are collected in the first row of this expression):

$$P_2(C_1) = (\alpha_0\beta_0\alpha_2\beta_2)^2 + (\alpha_0\beta_0\alpha_2\beta_3)^2 + (\alpha_0\beta_0\alpha_3\beta_2)^2 + (\alpha_0\beta_0\alpha_3\beta_3)^2 =$$

$$(\alpha_0\beta_0)^2 * \{(\alpha_2\beta_2)^2 + (\alpha_2\beta_3)^2 + (\alpha_3\beta_2)^2 + (\alpha_3\beta_3)^2\} = (\alpha_0\beta_0)^2 = P(C) \quad (13)$$

Here the sum in curly brackets is equal to 1 according to the normalization condition (11). The expression (13) means that the collective probability $P_2(C_1)$ of 4 doublets CC, CT, CG and CA is equal to the individual probability $P(C)$ of the letter C in this quantum-informational model of collective and individual probabilities inside long DNA-texts From the expressions (12) and (6), similar calculations of collective probabilities $P_2(T_1)$, $P_2(G_1)$ and $P_2(A_1)$ of all doublets with the first letters T, G, A in them give similar results (14) of their equality to individual probabilities of letters T, G and A:

$$P_2(T_1) = P(T), \quad P_2(G_1) = P(G), \quad P_2(A_1) = P(A) \quad (14)$$

These model results correspond to the phenomenological rule of symmetries formulated in the beginning of this Section.

The expression (12) allows a calculation of the collective probability $P_2(C_2)$ of all 4 doublets having the second letter C in them:

$$P_2(C_2) = (\alpha_0\beta_0\alpha_2\beta_2)^2 + (\alpha_0\beta_1\alpha_2\beta_2)^2 + (\alpha_1\beta_0\alpha_2\beta_2)^2 + (\alpha_1\beta_1\alpha_2\beta_2)^2 =$$

$$(\alpha_2\beta_2)^2 * \{(\alpha_0\beta_0)^2 + (\alpha_0\beta_1)^2 + (\alpha_1\beta_0)^2 + (\alpha_1\beta_1)^2\} = (\alpha_2\beta_2)^2 = P(C) \quad (15)$$

Here the sum in curly brackets is equal to 1 according to the normalization condition (7). The expression (15) means that the collective probability $P_2(C_2)$ of all doublets having the second letter C (that is, CC, TC, GC, AC) is also equal to the individual probability $P(C)$ of the letter C. From the expression (12), similar calculations of collective probabilities $P_2(T_2)$, $P_2(G_2)$ and $P_2(A_2)$ of doublets with the second letters T, G, A in them give similar results (16) of their equality to individual probabilities $P(T)$, $P(G)$ and $P(A)$ of letters T, G and A:

$$P_2(T_2) = P(T), \quad P_2(G_2) = P(G), \quad P_2(A_2) = P(A) \quad (16)$$

By analogy, to consider the case of 64 triplets one should accordingly expand the 16-dimensional Hilbert space $H_1 \otimes H_2 \otimes H_3 \otimes H_4$ to the 64-dimensional Hilbert space $H_1 \otimes H_2 \otimes H_3 \otimes H_4 \otimes H_5 \otimes H_6$, etc.

5 SOME CONCLUDING REMARKS

Returning to the Jordan's thoughts on quantum biology, the difference between biological and inanimate objects should be explained. Jordan correctly pointed out that inanimate objects were governed by the average random motion of millions of particles, such that the motion of a single molecule has no influence whatsoever on the whole object. This insight is usually credited to Erwin Schrödinger, who later claimed that life was different from inorganic chemistry because of its dependence on the dynamics of a small number of molecules. Jordan similarly argued that the few molecules that control the dynamics of living cells within the control center have a dictatorial influence, such that quantum-level events that govern their motion, such as Heisenberg's uncertainty principle, are amplified to influence the entire organism. Jordan believed that living organisms were uniquely able to carry out this amplification in a way that was conspicuously different from inanimate matter. In fact, Jordan had been thinking about this question for over a decade and had been using the term “Quantumbiologie” since the late 1930s. Jordan was convinced he could extend quantum indeterminism from the subatomic world to macroscopic biology. He even made a connection with free will by suggesting a link between quantum mechanics and psychology. Jordan's insistence that living organisms have a unique ability to amplify the quantum into the macroscopic world has a lot of resonance with modern views of quantum biology [McFadden, Al-Khalili, 2018].

The model approach described in this article belongs to the field of quantum biology and gives additional mathematical tools to develop quantum genetics using symmetrical properties of the DNA alphabets. The results presented and the mentioned rule of symmetries in long DNA sequences show additional evidences in favor of the idea of quantum computing in biological organisms discussed in works of many authors. Authors believe that phenomena of symmetry in molecular-genetic systems should be studied more and more deeply as very useful for developing biological sciences.

Acknowledgments

Some results of this paper have been possible due to a long-term cooperation between Russian and Hungarian Academies of Sciences on the topic “Non-linear models and symmetrologic analysis in biomechanics, bioinformatics, and the theory of self-organizing systems”, where S.V. Petoukhov was a scientific chief from the Russian Academy of Sciences. The authors are grateful to G. Darvas, E. Fimmel, M. He, Z.B. Hu, Yu.I. Manin and I.V. Stepanyan for their collaboration.

REFERENCES

Abbott D., Davies P.C.W. and Pati A.K., eds. (2008) *Quantum Aspects of Life*, foreword by Sir Roger Penrose, ISBN-13: 978-1-84816-253-2. https://doi.org/10.1142/9781848162556

Altaisky M.V., Filatov F.P., (2001) Genetic information and quantum gas, arXiv:quant-ph/0106123v1, submitted on 22.06.2001.

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

Chargaff E. (1971) Preface to a *Grammar of Biology: A hundred years of nucleic acid research, Science*, 172, p. 637-642. https://doi.org/10.1126/science.172.3984.637

Darvas G. (2018) Petoukhov's rules on symmetries in long DNA-texts, *Symmetry: Culture and Science*, 29, 2, 318-320 , https://doi.org/10.26830/symmetry_2018_2_318, http://journal-scs.symmetry.hu/abstract/?pid=673

Fickett J.W., Burks C. (1989) Development of a database for nucleotide sequences, In *Mathematical Methods for DNA Sequences* (Ed. Waterman M.S.), p. 1-34. Florida: CRC Press, Inc.

Fimmel, E., Danielli A., Strüngmann L. (2013) On dichotomic classes and bijections of the genetic code. *J. Theor. Biol*., *336*, 221–230. https://doi.org/10.1016/j.jtbi.2013.07.027

Fimmel E., Petoukhov S. (2019) Genetic Code Modeling from the Perspective of Quantum Informatics, In: *Advances in Artificial Systems for Medicine and Education II.* / Hu Z.B., He M., Petoukhov S.V. (Eds). Springer (in print), 978-3-030-12081-8, 478911_1_En. https://doi.org/10.1007/978-3-030-12082-5_11

Hu Z.B., Petoukhov S.V., Petukhova E.S. (2017a) Generalized crystallography, the genetic system and biochemical esthetics, *Structural Chemistry*, v. 28, №1, pp. 354-368. https://doi.org/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. (2017b) I-Ching, dyadic groups of binary numbers and the geno-logic coding in living bodies, *Progress in Biophysics and Molecular Biology,* 131, December, 354-368. https://doi.org/10.1016/j.pbiomolbio.2017.08.018 .

Hu Z.B., Petoukhov S.V., Petukhova E.S. (2018) On symmetries, resonances and photonic crystals in morphogenesis, *Biosystems*, online 14.09.2018, https://doi.org/10.1016/j.biosystems.2018.09.004

Igamberdiev A.I. (1993) Quantum mechanical properties of biosystems: a framework for complexity, structural stability, and transformations, *Biosystems*, 31 (1), 65–73. https://doi.org/10.1016/0303-2647(93)90018-8

Igamberdiev A.I. (2004) Quantum computation, non-demolition measurements, and reflective control in living systems, *BioSystems*, 77, 47–56. https://doi.org/10.1016/j.biosystems.2004.04.001

Igamberdiev A.I. (2007) Physical limits of computation and emergence of life, *BioSystems*, 90, 340–349. https://doi.org/10.1016/j.biosystems.2006.09.037

Igamberdiev A.I. (2008) Objective patterns in the evolving network of non-equivalent observers, *BioSystems*, 92, 122–131. https://doi.org/10.1016/j.biosystems.2008.01.002

Igamberdiev A.I., Shklovskiy-Kordi N.E. (2016) Computational power and generative capacity of genetic systems, *BioSystems*, 142–143, 1–8. https://doi.org/10.1016/j.biosystems.2016.01.003

Igamberdiev A.I., Shklovskiy-Kordi N.E. (2017) The quantum basis of spatiotemporality in perception and consciousness, *Progress in Biophysics and Molecular Biology*, 130, 15-25.

Jordan P. (1932) *Die Quantenmechanik und die Grundprobleme der Biologie und Psychologie. Naturwissenschaften* 20, 815–821. https://doi.org/10.1007/BF01494844 .

Josephson B.D. (2018) *The Physics of Mind and Thought,* Preprint. https://doi.org/10.13140/RG.2.2.36516.32640/2 , https://www.researchgate.net/publication/328968105 .

Matsuno K. (1999) Cell motility as an entangled quantum coherence, *BioSystems*, 51, 15–19. https://doi.org/10.1016/S0303-2647(99)00009-X

Matsuno K. (2003) Quantum mechanics in first, second and third person descriptions, *BioSystems*, 68, 107 118. https://doi.org/10.1016/S0303-2647(02)00090-4

Matsuno K., Paton R.C. (2000) Is there a biology of quantum information? *BioSystems*, 55, 39–46. https://doi.org/10.1016/S0303-2647(99)00081-7

McFadden J., Al-Khalili J. (2018) The origins of quantum biology, *Proceedings of the Royal Society A*: *Mathematical, Physical and Engineering Sciences*, 12 December, 1-13. https://doi.org/10.1098/rspa.2018.0674 https://royalsocietypublishing.org/doi/full/10.1098/rspa.2018.0674.

Mikheenko P. (2018) Possible superconductivity in brain. *https://arxiv.org/abs/1812.05602* , submitted on 13.12.2018. https://doi.org/10.1007/s10948-018-4965-4

Nielsen M.A., Chuang I.L. (2010) *Quantum Computation and Quantum Information,* New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511976667

Patel A. (2001a) Quantum algorithms and the genetic code, Pramana, *Journal of Physics*, 56, 2-3, 367-381, arXiv:quant-ph/0002037. https://doi.org/10.1007/s12043-001-0131-8

Patel A. (2001b) Testing quantum dynamics in genetic information processing, *Journal of Genetics*, 80, 1, 39 43. https://doi.org/10.1007/BF02811417

Patel A. (2001c) Why genetic information processing could have a quantum basis, *Journal of Biosciences*, 26, 2, 145-151. https://doi.org/10.1007/BF02703638

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

Penrose R. (2019) Your eyes are not meant for seeing, *Community*, February 21, https://thriveglobal.com/stories/your-eyes-are-not-meant-for-seeing/ .

Petoukhov S.V. (1999) Genetic Code and the Ancient Chinese "Book of Changes", *Symmetry: Culture and Science*, 10, 3-4, 211-226. https://doi.org/10.26830/symmetry_1999_3-4_211

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

Petoukhov S.V. (2016a) The system-resonance approach in modelling genetic structures, *Biosystems*, January, 139, 1-11. https://doi.org/10.1016/j.biosystems.2015.11.001, http://petoukhov.com/PETOUKHOV_ARTICLE_IN_BIOSYSTEMS.pdf .

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

Petoukhov S.V. (2018a) The Genetic Coding System and Unitary Matrices, *Preprints*, 2018040131. https://doi.org/10.20944/preprints201804.0131.v2

Petoukhov S.V. (2018b) The rules of long DNA-sequences and tetra-groups of oligonucleotides, arXiv:1709.04943v5, 5th version (8 October 2018), 159 pages.

Petoukhov S.V. (2019a) Structural Connections between Long Genetic and Literary Texts, *Preprints*, 2018120142 , online 15 02 2019. https://doi.org/10.20944/preprints201812.0142.v2 , https://www.preprints.org/manuscript/201812.0142/v2.

Petoukhov S.V. (2019b) Connections Between Long Genetic and Literary Texts; The Quantum-Algorithmic Modelling, In: Hu Z., Petoukhov S., Dychka I., He M. (eds) *Advances in Computer Science for Engineering and Education II*. pp 534-543, ICCSEEA; *Advances in Intelligent Systems and Computing*, vol 938, Springer, Cham. https://doi.org/10.1007/978-3-030-16621 -2_50, https://link.springer.com/chapter/10.1007/978-3-030-16621-2_50#citeas (online).

Petoukhov S.V., He M. (2009) *Symmetrical Analysis Techniques for Genetic Systems and Bioinformatics: Advanced Patterns and Applications*, Hershey, USA: IGI Global. 271 p. https://doi.org/10.4018/978-1-60566-124-7

Petoukhov S.V., Petukhova E.S., Svirin V.I. (2018) New Symmetries and Fractal-Like Structures in the Genetic Coding System, In: Hu Z., Petoukhov S., Dychka I., He M. (eds). *Advances in Computer Science for Engineering and Education*. ICCSEEA 2018, pp. 588-600; *Advances in Intelligent Systems and Computing*, vol 754, Springer, Cham, https://doi.org/10.1007/978-3-319-91008-6_59

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