BioSystems 209 (2021) 104503

Contents lists available at ScienceDirect

BioSystems

journal homepage: www.elsevier.com/locate/biosystems

Algebraic harmony and probabilities in genomes. Long-range coherence in quantum code biology

Sergey V. Petoukhov

Mechanical Engineering Research Institute of Russian Academy of Sciences, 101990, Moscow, M. Kharitonievskiy Pereulok, 4, Russia

ARTICLE INFO

Keywords:
Code biology
Genomes
Probability rules
Oligomer sums method
Matrices
Tensor product
Gestalt phenomena
Frohlich's theory
Quantum informatics

ABSTRACT

According to the founders of quantum mechanics and quantum biology P. Jordan and E. Schrödinger, the main difference between living and inanimate objects is the dictatorial influence of genetic molecules on the whole living organism. Code biology can make a valuable contribution to understanding this dictatorial influence of genetic molecules whose ensemble is endowed with many interconnected alphabets and codes. The paper is devoted to probability rules of nucleotide sequences of single-stranded DNA in eukaryotic and prokaryotic genomes. These rules are connected with *n*-plets alphabets of DNA whose nucleotide sequences are considered as bunches of many parallel texts written in interconnected *n*-plets alphabets. The rules draw attention to genomic phenomena of special tetragroupings of *n*-plets and new genomic symmetries. A generalization of the second Chargaff's rule is described. They show the existence of long-range coherence in genomic DNA sequences and reveal new connections of structural features of genomic sequences with formalisms of quantum mechanics and quantum informatics. The author supposes that the received results are related to the known vibration-resonance theory of G. Frohlich about long-range coherence in biological systems, that is, about collective quantum effects there. The possible influence of the described genetic probability phenomena on the genetically inherited physiological structures is noted and discussed.

1. Introduction

Science has led to a new understanding of life itself: «Life is a partnership between genes and mathematics» [Stewart, 1999]. But what kind of mathematics is a partner with the genetic code? Trying to find such mathematics, the author has turned to study the multi-level system of interrelated molecular-genetic alphabets. Alphabets play a basic role in communication technologies. In any communication system of "transmitter-receiver", the receiver always knows the alphabet of signals, which are used by the transmitter.

It is known that the molecular-genetic system of living bodies includes, in particular, the following alphabets, each of which can be considered as a part of a complex alphabetic system: the 4-letter alphabet of nitrogenous bases; the 64-letter alphabet of triplets; the 2-letter alphabet of "weak and strong roots" of triplets; the 20-letter alphabet of amino acids; the 2-letter alphabet "purines vs. pyrimidines"; the 2-letter alphabet "strong vs. weak hydrogen bonds''; the 2-letter alphabet "keto vs. amino", etc. (see a wide list of genetic alphabets in [Karlin et al., 1989]). A profound study of the phenomenological fact of the parallel existence of a wide set of different and interconnected alphabets and codes in genetic informatics is important for developing the code biology [Barbieri, 2015] and understanding living bodies as genetically inherited holistic essences.

According to the founders of quantum mechanics and quantum biology P. Jordan and E. Schrödinger, the main difference between living and inanimate objects: inanimate objects are controlled by the average random movement of their millions of particles, whose individual influence is negligible, while in a living organism selected – genetic - molecules have a dictatorial influence on the whole living organism. Besides this, Jordan 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 percentages of *n*-plets (monoplets, doublets, triplets, etc., that is, oligomers with lengths n) in long DNA sequences is important for discovering hidden biological laws and for developing quantum biology. Correspondingly, this article is devoted to the author's results of studying hidden rules of probabilities in long nucleotide sequences of single-stranded DNA in eukaryotic and prokaryotic genomes.

In his previous articles, the author described the universal hyperbolic

E-mail addresses: spetoukhov@gmail.com, spetoukhov@gmail.com.

https://doi.org/10.1016/j.biosystems.2021.104503
Received 29 April 2021; Received in revised form 4 August 2021; Accepted 4 August 2021
Available online 20 August 2021

rules of the oligomer cooperative organization of DNA nucleotide sequences in eukaryotic and prokaryotic genomes [Petoukhov, 2020a-c]. These rules were discovered based on a new method of oligomer sums for the analysis of long nucleotide sequences. Let us remind this method since it is also used in this article and is connected with interrelated DNA alphabets of *n*-plets. As it is known, there are DNA alphabets of 4 nucleotides (adenine A, cytosine C, guanine G, thymine T), 16 doublets, 64 triplets, etc. (each such alphabet of *n*-plets consists of 4^n elements of length *n*). The mentioned method represents long DNA sequences in the form of composite, multilayered texts, in which each *n*th layer is a sequence of *n*-plets (or oligomers of fixed length *n*); in other words, each *n*th layer is a separate DNA-text written in its alphabet of 4^n *n*-plets. For example, in the sequence ACCTGTAACG …, the first layer is a text, which is based on the alphabet of 4 nucleotides (A-C-C-C-T-G- …), the second layer is a text, which is based on the alphabet of 16 doublets (AC-CT-GT-AA-CG- …), the third is a text, which is based on the alphabet of 64 triplets (ACC-TGT-AAC- …), etc. In each *n*th layer, one calculates the percentages of each of the 4^n members of the *n*-plets alphabet and studies the relationship between all their percentages in different layers. This approach to the analysis of any genomic sequence as a set of many parallel texts, each of which is written in its alphabet (but all these alphabets are interrelated), reveals important algebraic patterns in the genetic informatics of higher and lower organisms.

For comfortable showing the calculated set of phenomenological percentages of *n*-plets in genomic DNA layers, one uses a family of square tables, which contain complete sets of DNA alphabets of *n*-plets and was early described in [Petoukhov, 2008; Petoukhov, He, 2010]. These tables (Fig. 1) are based on the system of binary-oppositional molecular traits in the DNA alphabet of four nucleotides C, A, T, and G:

1) two of these molecules are purines with two rings (A and G), and the other two are pyrimidines with one ring (C and T). In terms of these oppositional indicators, one can represent C = T = 1, A = G = 0;
2) the two letters are keto molecules (T and G), and the other two - amino molecules (C and A). In terms of these oppositional indicators, ony can represent C = A = 1, T = G = 0.

In each of the alphabetic tables, its columns are enumerated in accordance with the oppositional indicators «purine or pyrimidine» (C = T = 1, A = G = 0), and its rows are enumerated in accordance with oppositional indicators «keto or amino» (C = A = 1, T = G = 0). In such tables, all monoplets, doublets, triplets and other *n*-plets automatically occupy their strictly individual places (see examples of the alphabetic tables in Fig. 1). This system of alphabetic tables coincides with a tensor family of matrices $[C, A; T, G]^{(n)}$ where (n) refers to tensor power *n* [Petoukhov, 2008; Petoukhov, He, 2010]. Correspondingly, each of these alphabetic tables of *n*-plets can be interpreted as an appropriate alphabetic matrix $[C, A; T, G]^{(n)}$. Appendix I describes the properties of the tensor product of matrices and its important role in quantum mechanics and quantum informatics.

The article aims to represent and discuss the results of the analysis of possible interrelations of percentages of different *n*-plets in text layers of DNA sequences of eukaryotic and prokaryotic genomes.

2. Positional tetra-groupings rule of percentage composition of *n*-plets

For the study, nucleotide sequences in single-stranded DNA of eukaryotic and prokaryotic genomes are taken from the well-known

	1	0
1	C	A
0	T	G

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

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

	1111	1110	1101	1100	1011	1010	1001	1000	0111	0110	0101	0100	0011	0010	0001	0000
1111	CCCC	CCCA	CCAC	CCAA	CACC	CACA	CAAC	CAAA	ACCC	ACCA	ACAC	ACAA	AACC	AACA	AAAC	AAAA
1110	CCCT	CCCG	CCAT	CCAG	CACT	CACG	CAAT	CAAG	ACCT	ACCG	ACAT	ACAG	AACT	AACG	AAAT	AAAG
1101	CCTC	CCTA	CCGC	CCGA	CATC	CATA	CAGC	CAGA	ACTC	ACTA	ACGC	ACGA	AATC	AATA	AAGC	AAGA
1100	CCTT	CCTG	CCGT	CCGG	CATT	CATG	CAGT	CAGG	ACTT	ACTG	ACGT	ACGG	AATT	AATG	AAGT	AAGG
1011	CTCC	CTCA	CTAC	CTAA	CGCC	CGCA	CGAC	CGAA	ATCC	ATCA	ATAC	ATAA	AGCC	AGCA	AGAC	AGAA
1010	CTCT	CTCG	CTAT	CTAG	CGCT	CGCG	CGAT	CGAG	ATCT	ATCG	ATAT	ATAG	AGCT	AGCG	AGAT	AGAG
1001	CTTC	CTTA	CTGC	CTGA	CGTC	CGTA	CGGC	CGGA	ATTC	ATTA	ATGC	ATGA	AGTC	AGTA	AGGC	AGGA
1000	CTTT	CTTG	CTGT	CTGG	CGTT	CGTG	CGGT	CGGG	ATTT	ATTG	ATGT	ATGG	AGTT	AGTG	AGGT	AGGG
0111	TCCC	TCCA	TCAC	TCAA	TACC	TACA	TAAC	TAAA	GCCC	GCCA	GCAC	GCAA	GACC	GACA	GAAC	GAAA
0110	TCCT	TCCG	TCAT	TCAG	TACT	TACG	TAAT	TAAG	GCCT	GCCG	GCAT	GCAG	GACT	GACG	GAAT	GAAG
0101	TCTC	TCTA	TCGC	TCGA	TATC	TATA	TAGC	TAGA	GCTC	GCTA	GCGC	GCGA	GATC	GATA	GAGC	GAGA
0100	TCTT	TCTG	TCGT	TCGG	TATT	TATG	TAGT	TAGG	GCTT	GCTG	GCGT	GCGG	GATT	GATG	GAGT	GAGG
0011	TTCC	TTCA	TTAC	TTAA	TGCC	TGCA	TGAC	TGAA	GTCC	GTCA	GTAC	GTAA	GGCC	GGCA	GGAC	GGAA
0010	TTCT	TTCG	TTAT	TTAG	TGCT	TGCG	TGAT	TGAG	GTCT	GTCG	GTAT	GTAG	GGCT	GGCG	GGAT	GGAG
0001	TTTC	TTTA	TTGC	TTGA	TGTC	TGTA	TGGC	TGGA	GTTC	GTTA	GTGC	GTGA	GGTC	GGTA	GGGC	GGGA
0000	TTTT	TTTG	TTGT	TTGG	TGTT	TGTG	TGGT	TGGG	GTTT	GTTG	GTGT	GTGG	GGTT	GGTG	GGGT	GGGG

Fig. 1. The tabular representation of the DNA-alphabets of 4 nucleotides, 16 doublets, 64 triplets, and 256 tetraplets. These tables are constructed by the described binary numbering of their rows and columns, and they coincide with appropriate alphabetic matrices $[C, A; T, G]^{(n)}$ where (n) refers to the tensor power *n*.

Genbank. In each of the studied layers of genomic DNA, an individual quantity of each of the appropriate 4^n members of the *n*-plets alphabet is calculated; then this quantity is divided by the total amount of all *n*-plets in this layer to define a percent of this kind of the *n*-plets in the layer. One can remind that genomic sequences in the GenBank sites usually contain some letters N, which indicate that there can be any nucleotide in this place (https://www.ncbi.nlm.nih.gov/books/NBK21136/). For this reason, the total amount of all nucleotides A, T, C, G, which are calculated for the sequence from the GenBank, is slightly less than the complete length of the DNA sequence, which is indicated in the GenBank. But practically this is not essential for the resulting values of percentages of separate nucleotides in the analyzed genomic sequences.

To explain general obtained results, let us use a particular example of the DNA of the human chromosome N♌1, which contains a sequence of about 250 million nucleotides C, A, T, and G (initial data on this chromosome were taken in the GenBank: https://www.ncbi.nlm.nih.gov/nuccore/NC_000001.11). At the first step of the analysis, the percents of each of the nucleotides C, A, T, and G in this chromosome are calculated: %C ≈ 0.2085, %G ≈ 0.2089, %A ≈ 0.2910, %T ≈ 0.2917 (here percents are shown in fractions of one). Here and below, percentages are usually rounded to the fourth decimal place. These percent values are used to be indicated in appropriate cells of the matrix of nucleotides (Fig. 1) instead of nucleotide symbols for receiving a numeric matrix of nucleotides percents (Fig. 2, upper row). One can note that here %C ≈ %G and %A ≈ %T in line with the second Chargaff's rule [Albrecht-Buehler, 2006; Chargaff, 1971; Prabhu, 1993].

At the second step of the described approach, the nucleotide sequence of the analyzed chromosome is represented as a text of doublets, in which the percentage of each of 16 doublets is calculated. Then all these percents are indicated in appropriate cells of the (4*4)-matrix (Fig. 2, bottom row). At the third step of the described approach, the nucleotide sequence of the analyzed chromosome is represented as a text of triplets, in which the percentage of each of 64 triplets is calculated. Then these percents are indicated in cells of the (8*8)-matrix [C, A; T, G]$^{(3)}$ (Fig. 3).

At the fourth step of the described approach, the DNA sequence of the analyzed chromosome is represented as a text of tetraplets, and percents of each of 256 tetraplets are calculated. Then these percents are indicated in appropriate cells of the (16*16)-matrix [C, A; T, G]$^{(4)}$ shown in Fig. 1. The resulting matrix of percent of 256 tetraplets is presented in Fig. 4.

At first glance, the set of percent in the resulting alphabetic matrices (Figs. 2–4) is quite chaotic. It has the following features regarding the percent of separate *n*-plets:

- Percent of presented *n*-plets significantly depends on the order of letters in them. For example, the percent of doublets CG and GC, having the same letter composition, differ several times: %CG = 0.0103, and %GC = 0.0440. Similarly, the percent of triplets of the same letter composition CAT, CTA, ACT, ATC, TCA, TAC are significantly different: %CAT = 0.0179, %CTA = 0.0127, %ACT = 0.0162, %ATC = 0.0132, %TCA = 0.0196, %TAC = 0.0110, and so on;
- Accordingly, the matrices of phenomenological percentages of doublets, triplets, and tetraplets (Figs. 2–4) don't coincide numerically with matrices of the tensor family [%C, %A; %T, %G]$^{(n)}$ = [0.2085, 0.2910; 0.2918, 0.2087]$^{(n)}$.

But unexpectedly these values %C = 0.2085, %G = 0.2087, %A = 0.2910, %T = 0.2917 show themselves in the block organization of percentages of different *n*-plets in various layers of the genomic DNA-text as one can calculate from data of percent matrices in Figs. 2–4. For example, the following sums of percentages of *n*-plets, which have the nucleotide C as an attributive indicator of their grouping, are realized:

- The total sum Σ%CN of percentages of all 4 doublets CN (hereinafter, the symbol N denotes any of the nucleotides A, T, C, and G), which start with the nucleotide C, is equal to %C, that is, Σ%CN ≈ %CC + %CA +%CT + %CG ≈ 0.0541 + 0.0727+0.0713 + 0.0103 ≈ 0.2085 ≈ %C;
- The total sum Σ%NC of percentages of all 4 doublets NC, which have the nucleotide C at their second positions, is also practically equal to %C, that is, Σ%NC ≈ %CC +%AC + %TC +%GC = 0.0541 + 0.0503+0.0601 + 0.0440 ≈ 0.2085 ≈ %C;
- The total sum Σ%CNN of percentages of all 16 triplets CNN, which have the nucleotide C at their first position, is also practically equal to %C, that is, Σ%CNN ≈ 0.0284 ≈ %C;
- The total sum Σ%NCN of percentages of all 16 triplets NCN, which have the nucleotide C at their second position, is also practically equal to %C, that is, Σ%NCN ≈ 0.0285 ≈ %C;
- The total sum Σ%NNC of percentages of all 16 triplets NNC, which have the nucleotide C at their third position, is also practically equal to %C, that is, Σ%NNC ≈ 0.0285 ≈ %C;
- The total sum Σ%CNNN of percentages of all 64 tetraplets CNNN, which have the nucleotide C at their first position, is also practically equal to %C, that is, Σ%CNNN ≈ 0.0285 ≈ %C;
- The total sum Σ%NCNN of percentages of all 64 tetraplets NCNN, which have the nucleotide C at their second position, is also practically equal to %C, that is, Σ%NCNN ≈ 0.0285 ≈ %C;
- The total sum Σ%NNCN of percentages of all 64 tetraplets NNCN, which have the nucleotide C at their third position, is also practically equal to %C, that is, Σ%NNCN ≈ 0.0285 ≈ %C;

C	A
T	G

➔

%C	%A
%T	%G

=

0.2085	0.2910
0.2918	0.2087

%CC	%CA	%AC	%AA
%CT	%CG	%AT	%AG
%TC	%TA	%GC	%GA
%TT	%TG	%GT	%GG

=

0.05409	0.07274	0.05033	0.09504
0.07134	0.01031	0.07429	0.07137
0.06008	0.06312	0.04402	0.06008
0.09568	0.07286	0.05046	0.05419

Fig. 2. The transformation of the symbolic matrices of 4 nucleotides and 16 doublets from Fig. 1 into appropriate numeric matrices of the percentage of nucleotides and doublets in the case of the human chromosome N♌1.

%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

=

0.0138	0.0188	0.0152	0.0186	0.0118	0.0198	0.0145	0.0369
0.0185	0.0029	0.0179	0.0210	0.0162	0.0025	0.0238	0.0199
0.0176	0.0127	0.0025	0.0023	0.0132	0.0194	0.0144	0.0224
0.0201	0.0209	0.0026	0.0029	0.0239	0.0178	0.0161	0.0185
0.0159	0.0196	0.0110	0.0199	0.0125	0.0146	0.0096	0.0196
0.0223	0.0023	0.0194	0.0128	0.0144	0.0025	0.0133	0.0176
0.0197	0.0198	0.0146	0.0195	0.0096	0.0112	0.0126	0.0160
0.0372	0.0188	0.0199	0.0190	0.0145	0.0153	0.0119	0.0138

Fig. 3. The matrix of percent of the 64 triplets in the DNA-sequence of triplets in the human chromosome №1.

.0033	.0055	.0042	.0044	.0040	.0056	.0032	.0070	.0030	.0042	.0040	.0053	.0032	.0059	.0055	.0149
.0041	.0010	.0044	.0058	.0047	.0010	.0040	.0044	.0041	.0005	.0051	.0054	.0047	.0006	.0095	.0071
.0050	.0029	.0008	.0006	.0036	.0039	.0049	.0059	.0037	.0032	.0006	.0006	.0040	.0070	.0037	.0066
.0048	.0058	.0006	.0009	.0057	.0047	.0045	.0058	.0049	.0044	.0007	.0007	.0071	.0057	.0049	.0047
.0052	.0057	.0027	.0038	.0010	.0006	.0003	.0006	.0033	.0045	.0032	.0063	.0042	.0049	.0039	.0078
.0058	.0009	.0035	.0028	.0007	.0003	.0005	.0008	.0049	.0005	.0064	.0035	.0046	.0007	.0049	.0059
.0048	.0036	.0046	.0051	.0005	.0004	.0008	.0006	.0047	.0056	.0033	.0050	.0031	.0037	.0047	.0057
.0073	.0044	.0053	.0058	.0006	.0010	.0005	.0010	.0096	.0040	.0051	.0044	.0046	.0047	.0040	.0041
.0046	.0049	.0042	.0051	.0024	.0045	.0028	.0078	.0030	.0041	.0029	.0038	.0023	.0037	.0029	.0073
.0057	.0006	.0051	.0052	.0037	.0004	.0056	.0036	.0047	.0008	.0033	.0046	.0030	.0005	.0046	.0048
.0057	.0040	.0005	.0006	.0030	.0055	.0026	.0041	.0032	.0027	.0006	.0005	.0025	.0031	.0032	.0057
.0068	.0058	.0006	.0006	.0070	.0039	.0032	.0029	.0037	.0048	.0006	.0008	.0041	.0036	.0036	.0051
.0051	.0063	.0034	.0063	.0041	.0049	.0031	.0061	.0023	.0031	.0017	.0035	.0031	.0042	.0023	.0052
.0077	.0006	.0063	.0038	.0049	.0006	.0045	.0057	.0038	.0004	.0032	.0027	.0042	.0010	.0034	.0052
.0073	.0078	.0038	.0051	.0037	.0046	.0042	.0051	.0030	.0028	.0029	.0042	.0023	.0024	.0030	.0046
.0150	.0072	.0055	.0045	.0059	.0057	.0043	.0054	.0055	.0032	.0040	.0042	.0032	.0039	.0030	.0033

Fig. 4. The matrix of percents of the 256 tetraplets in the DNA-sequence of tetraplets in the human chromosome №1.

%C ≈ 0.2085	%G ≈ 0.2087	%A ≈ 0.2910	%T ≈ 0.2918
Σ%CN ≈ 0.2085	Σ%GN ≈ 0.2088	Σ%AN ≈ 0.2910	Σ%TN ≈ 0.2917
Σ%NC ≈ 0.2085	Σ%NG ≈ 0.2087	Σ%NA ≈ 0.2910	Σ%NT ≈ 0.2918
Σ%CNN ≈ 0.2084	Σ%GNN ≈ 0.2088	Σ%ANN ≈ 0.2910	Σ%TNN ≈ 0.2917
Σ%NCN ≈ 0.2085	Σ%NGN ≈ 0.2088	Σ%NAN ≈ 0.2910	Σ%NTN ≈ 0.2917
Σ%NNC ≈ 0.2085	Σ%NNG ≈ 0.2087	Σ%NNA ≈ 0.2910	Σ%NNT ≈ 0.2918
Σ%CNNN ≈ 0.2085	Σ%GNNN ≈ 0.2088	Σ%ANNN ≈ 0.2910	Σ%TNNN ≈ 0.2917
Σ%NCNN ≈ 0.2085	Σ%NGNN ≈ 0.2087	Σ%NANN ≈ 0.2910	Σ%NTNN ≈ 0.2918
Σ%NNCN ≈ 0.2085	Σ%NNGN ≈ 0.2088	Σ%NNAN ≈ 0.2910	Σ%NNTN ≈ 0.2918
Σ%NNNC ≈ 0.2085	Σ%NNNG ≈ 0.2087	Σ%NNNA ≈ 0.2910	Σ%NNNT ≈ 0.2918

Fig. 5. Percentages of nucleotides C, G, A, T (in the first row), and the sum Σ of percents of *n*-plets with these nucleotides at their certain positions in the case of the first 4 layers of the DNA-sequence of the human chromosome №1 (here $n = 1, 2, 3, 4$). The symbol N denotes any of the nucleotides.

- The total sum Σ%NNNC of percentages of all 64 tetraplets NNNC, which have the nucleotide C at their fourth position, is also practically equal to %C, that is, Σ%NNCN ≈ 0.0285 ≈ %C.

A similar phenomenological situation holds for other groupings of *n*-plets, for which other nucleotides G, A, T play roles of the attributive indicators. Fig. 5 shows this stochastic phenomenon of the constant values of sums of percentages of appropriate members of *n*-plets alphabets in *n*-plets layers of the DNA-sequence of the human chromosome N⁰1.

Briefly speaking, the following equalities (2.1) hold - with a high level of accuracy - regarding interrelated percentages in the considered 4 groupings of *n*-plets from different layers of the DNA-text of the human chromosome N⁰1:

$$\%C \approx \sum \%CN \approx \sum \%NC \approx \sum \%CNN \approx \sum \%NCN \approx \sum \%NNC \approx \sum \%CNNN \approx \sum \%NCNN \approx \sum \%NNCN \approx \sum \%NNNC$$
$$\%G \approx \sum \%GN \approx \sum \%NG \approx \sum \%GNN \approx \sum \%NGN \approx \sum \%NNG \approx \sum \%GNNN \approx \sum \%NGNN \approx \sum \%NNGN \approx \sum \%NNNG$$
$$\%A \approx \sum \%AN \approx \sum \%NA \approx \sum \%ANN \approx \sum \%NAN \approx \sum \%NNA \approx \sum \%ANNN \approx \sum \%NANN \approx \sum \%NNAN \approx \sum \%NNNA$$
$$\%T \approx \sum \%TN \approx \sum \%NT \approx \sum \%TNN \approx \sum \%NTN \approx \sum \%NNT \approx \sum \%TNNN \approx \sum \%NTNN \approx \sum \%NNTN \approx \sum \%NNNT \quad (2.1)$$

Knowing the percentages of nucleotides %A, %T, %C, and %G, it is possible to predict with high accuracy the sums of percentages of *n*-plets from the appropriate C-, G-, A-, T-groupings represented inside columns in Fig. 5. The ability of such predictions based on equalities (2.1) exists not only for the considered human chromosome N⁰1 but also for all eukaryotic and prokaryotic genomes, which were analyzed by the author till now including the following:

- 1) all 24 human chromosomes, which differ in their length, the number, and type of genes, etc.;
- 2) all chromosomes of a fruit fly *Drosophila melanogaster*, all chromosomes of a house mouse *Mus musculus*, all chromosomes of a nematode *Caenorhabditis elegans*, all chromosomes of a plant *Arabidopsis thaliana*, and many other plants;
- 3) 19 bacterial genomes of different groups both from Bacteria and Archaea.

One should add that percentages of nucleotides %A, %T, %C, and %G can be essentially different in various genomes. For example, the genomic DNA of bacteria *Bradyrhizobium japonicum* has %A≈ 0.1819, %T≈0.1815, %C≈0.3184, and %G≈0.3182 in contrast to the considered case of the human chromosome N⁰1 (see percentage data of this bacterial genome in [Petoukhov, 2021b, Appendix I]).

The four columns in Fig. 5 show that in each of the presented layers of the genomic DNA-text, there exist alphabetic tetra-groupings of *n*-plets with the same percentage sums. The following **positional rule of probabilities** in layers of long DNA-sequences, which is a candidacy for the universal rule for genomes, can be formulated based on such results:

- All *n* layers of a long DNA-sequence, each of which consists of a text of 4^n *n*-plets, has approximately the same sum of percentages of all those *n*-plets, which contain the considered nucleotide (C, G, A, or T) at a fixed *m*th position ($m \leq n$); here $n = 1, 2, 3, 4, \ldots$ but is not too large compared to the length of DNA.

By this rule, the percentage sums of *n*-plets in such alphabetic groupings in different layers of a considered long DNA-sequence are equal to the percentage of a corresponding nucleotide (that is, %C, %G, %A, or %T) in the first layer of the DNA-sequence, although the percent values of individual *n*-plets, that are summands in these sums, can differ significantly. For example, in the second layer of the DNA-sequence of the human chromosome N⁰ 1, the sum of the percentages of all 4 doublets with nucleotide C in their first position ($m = 1$) is equal to Σ%CN ≈ %CC+%CA+%CT+%CG ≈ 0.0541 + 0.0727 + 0.0713 + 0.0103 ≈ 0.2085. In the same second layer, the sum of the percentages of all 4 doublets with a nucleotide C in the second position ($m = 2$) is equal to the same number, although the summands in this sum are significantly different: Σ%NC ≈ %CC+%AC+%TC+%GC = 0.0541 + 0.0503 + 0.0601 + 0.0440 ≈ 0.2085. These two equal total values are equal to the percentage of nucleotide C in the first layer of the given genomic DNA-sequence: %C ≈ 0.0285.

The formulated positional rule for multi-layer DNA-sequences is a generalization of the second Chargaff's rule speaking, as known, that %C ≈ %G and %A ≈ %T in long single-stranded DNA sequences (this Chargaff's rule speaks about probabilities only in the first layer of long DNA-sequences without analyzing probabilities of *n*-plets in other layers). The generalized Chargaff rule cannot be analytically derived from the original Chargaff rule. Let us explain this by the example of comparing the percentage compositions of 4 nucleotides in the first layer of the genomic DNA sequence and 16 doublets in its second layer. With a fixed percentage of 4 nucleotides in the first layer of the DNA sequence, there are many options for the sequence of these nucleotides one after another. Each of these variants corresponds to an individual variant of the sequence of 16 doublets and also the number of doublets of each kinds in the second layer of the DNA sequence. In other words, with the same percentage composition of 4 nucleotides in the first layer of genomic DNA, there are many variants of the percentage composition of 16 doublets in its second layer. It follows that the rule of percentage composition of 16 doublets in the second layer cannot be analytically deduced from the percentage composition of 4 nucleotides in the first layer of this DNA. The same applies to the percentages of *n*-plets in the corresponding *n* layers of genomic DNA sequences.

Each of the tetra-groupings of *n*-plets, which is defined by a disposition of attributive nucleotides C, G, A, and T at a certain position *m* ($m \leq n$) in these *n*-plets, we call an *m*-positional tetra-grouping. For example, a tetra-grouping corresponding to sets ANN, TNN, CNN, and GNN is called a 1-positional tetra-grouping; a tetra-grouping corresponding to NAN, NTN, NCN, NGN is called a 2-positional tetra-grouping, and so on. The described stochastic phenomena have some analogies with the phenomena of holography, in which it is possible to reconstruct the image of a whole object from the image of its piece. Indeed, the knowledge of the sums of the percentages of *n*-plets in *m*-positional tetra-groupings of one layer of a long DNA-text gives knowledge about the sums of the percentages of *n*-plets in *m*-positional tetra-groupings of other layers.

Returning for a moment to the tensor family of matrices [C, A; T, G]$^{(n)}$ (Fig. 1), let us consider model (or reference) percentages of *n*-plets in a tensor family of percentage matrices [%C, %A; %T, %G]$^{(n)}$ ≈ [0.2085, 0.2910; 0.2917, 0.2089]$^{(n)}$. Figs. 6 and 7 show the received matrices for $(n) = 2, 3$, whose percent entries are significantly differ from real percentages of *n*-plets shown above in Figs. 2–4.

But unexpectedly the Gestalt rule holds for these model percentages of *n*-plets by the expression (2.2) as well:

$$
\begin{aligned}
&[\%A,\%T,\%C,\%G] \approx [\ \%AN,\ \%TN,\ \%CN,\ \%GN] \approx \\
&[\textstyle\sum \%NA, \sum \%NT, \sum \%NC, \sum \%NG] \approx \\
&[\textstyle\sum \%ANN, \sum \%TNN, \sum \%CNN, \sum \%GNN] \approx \\
&[\textstyle\sum \%NAN, \sum \%NTN, \sum \%NCN, \sum \%NGN] \approx \\
&[\textstyle\sum \%NNA, \sum \%NNT, \sum \%NNC, \sum \%NNG] \approx [0.2910,\ 0.2918,\ 0.2085,\ 0.2087]
\end{aligned}
\tag{2.2}
$$

It should be especially noted that in the case of a tensor family of percentage matrices $[\%C, \%A; \%T, \%G]^{(n)}$ we have the absolute accuracy of fulfillment of the Gestalt rule. For example, if %A = 0.291001313, %T = 0.291755765, %C = 0.208498924, %G = 0.208743998, then the expressions (2.1) is fulfilled precisely:

$$
\begin{aligned}
&[\%A,\%T,\%C,\%G] = [\ \textstyle\sum \%AN, \sum \%TN, \sum \%CN, \sum \%GN] = \\
&[\textstyle\sum \%NA, \sum \%NT, \sum \%NC, \sum \%NG] = \\
&[\textstyle\sum \%ANN, \sum \%TNN, \sum \%CNN, \sum \%GNN] = \\
&[\textstyle\sum \%NAN, \sum \%NTN, \sum \%NCN, \sum \%NGN] = \\
&[\textstyle\sum \%NNA, \sum \%NNT, \sum \%NNC, \sum \%NNG] = \\
&[\textstyle\sum \%ANNN, \sum \%TNNN, \sum \%CNNN, \sum \%GNNN] = \\
&[\textstyle\sum \%NANN, \sum \%NTNN, \sum \%NCNN, \sum \%NGNN] = \\
&[\textstyle\sum \%NNAN, \sum \%NNTN, \sum \%NNCN, \sum \%NNGN] = \\
&[\textstyle\sum \%NNNA, \sum \%NNNT, \sum \%NNNC, \sum \%NNNG] = \\
&[0.291001313,\ 0.291755765,\ 0.208498924,\ 0.208743998]
\end{aligned}
\tag{2.3}
$$

It gives pieces of evidence that fundamental genetic phenomena, reflected in the formulated Gestalt rule, are connected with the algebraic operation of the tensor product, which is so important in quantum mechanics and quantum informatics. Accordingly, the tensor family of percentage matrices $[\%C, \%A; \%T, \%G]^{(n)}$ and their percentage entries can be considered in each specific case as a certain standard of comparison in the analysis of long DNA-texts. Below it will show that the difference between the tetra-groupings of real percentages and these reference percentages is related to unitary operators.

The described *m*-positional tetra-groupings of *n*-plets can be transformed into each other based on cyclic shifts of positions in *n*-plets. Such transformations of positions in *n*-plets have accompanied a reconstruction of the (2^n*2^n)-matrices of *n*-plets alphabets (Fig. 1), where each of 4 matrix quadrants will contain in each time a complete grouping of *n*-plets with the same attributive nucleotide in their appropriate position *m*. Arrangements of *n*-plets, which are belonged to C-, G-, A-, T-groupings, in such matrices have some connection with algebras of hypercomplex numbers in matrix forms of their presentations as it is described in the preprint [Petoukhov, 2021b].

The genetic matrices of *n*-plets alphabets (Fig. 1) were constructed based on binary-oppositional traits of the DNA alphabet of 4 nucleotides. The genetic coding system has binary-oppositional structures at different levels of its organization. As it is known, the ancient Chinese book I-Ching, which was written a few thousand years ago, has introduced the system of symbols Yin and Yang (equivalents of 0 and 1). This book had a powerful impact on the culture, medicine, and science of ancient China and several other countries. The system of I-Ching is represented by the schemes with 4 bigrams, 8 trigrams, and 64 hexagrams. Similar to this, the genetic code is constructed on DNA molecules using 4 nitrogenous bases, 16 doublets, and 64 triplets. Structural analogies of Yin-Yang schemes of I-Ching with the alphabets of DNA have long been noted by various authors, including prominent geneticists: Stent G.S. in 1969 and Nobel laureate F. Jacob in 1977. The Ancient Chinese claimed that the table of 64 hexagrams in Fu-Xi's order is the universal archetype of nature. The Ancient Chinese knew nothing about the genetic code, but the genetic code is arranged following the I-Ching in many aspects. In particular, the genetic matrices of *n*-plets alphabets (Fig. 1) with binary numerations of their columns and rows are constructed by a deep analogy with a historical famous table of 64 hexagrams in Fu-Xi's order. More details about such analogies one can read in publications [Petoukhov, 1999, 2008; Petoukhov, He, 2010; Hu et al., 2017; Shchutskii, 1997].

I-Ching, which is called also «The Book of Cyclic Changes», declares a universality of the cyclic principle of organization in nature. Traditional Oriental medicine is based on the viewpoints of this book. Many western scientists studied and used I-Ching. For example, the creator of analytical psychology C. Jung developed his doctrine about the collective unconscious in connection with this book. According to Jung and his fellow campaigner Nobel laureate in physics W. Pauli, the trigrams and the hexagrams of I-Ching "*fix a universal set of archetypes (innate psychic structures)*" [Shchutskii, 1997, p. 12]. Many modern physicists, who feel the unity of the world, connect their theories with the ideas of traditional Oriental culture, which unite all nature. The intensive development of the self-organizing and nonlinear dynamics of complex systems promotes the strengthening attention of western scientists to the traditional eastern worldview (e.g., see [Capra, 2010]).

One should note that the genetic matrix $[C, A; T, G]^{(3)}$ of 64 triplets (Fig. 1), which are distributed among three different *m*-positional C-, G-, A-, and G-groupings (here $m = 1, 2, 3$), and the table of 64 hexagrams from I-Ching have deep structural analogies, which were unknown previously and which are described in the preprint [Petoukhov, 2021b]. Correspondingly, presented phenomenological data about stochastic regularities in genomic DNA-sequences can be considered in their connection with a rich theme of nature's archetypes and genetically inherited psychological phenomena including phenomena of Gestalt psychology, which are concerned below.

%CC	%CA	%AC	%AA
%CT	%CG	%AT	%AG
%TC	%TA	%GC	%GA
%TT	%TG	%GT	%GG

=

0.0435	0.0607	0.0607	0.0847
0.0608	0.0435	0.0849	0.0607
0.0608	0.0849	0.0435	0.0607
0.0851	0.0609	0.0609	0.0436

Fig. 6. The matrix of model percents for 16 doublets, which is the second tensor power of the matrix [0.2085, 0.2910; 0.2917, 0.2089] in the case of the human chromosome №1 (all values are rounded to the fourth decimal place).

%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

=

.00906382	.01265035	.01265035	.01765605	.01265035	.01765605	.01765605	.02464250
.01268315	.00907448	.01770183	.01266522	.01770183	.01266522	.02470639	.01767681
.01268315	.01770183	.00907448	.01266522	.01770183	.02470639	.01266522	.01767681
.01774773	.01269806	.01269806	.00908514	.02477045	.01772264	.01772264	.01268011
.01268315	.01770183	.01770183	.02470639	.00907448	.01266522	.01266522	.01767681
.01774773	.01269806	.02477045	.01772264	.01269806	.00908514	.01772264	.01268011
.01774773	.02477045	.01269806	.01772264	.01269806	.01772264	.00908514	.01268011
.02483467	.01776859	.01776859	.01271298	.01776859	.01271298	.01271298	.00909582

Fig. 7. The matrix of conditional or model percents for 64 triplets, which is the third tensor power of the matrix [0.2085, 0.2910; 0.2917, 0.2089] in the case of the human chromosome N☺1.

3. DNA epi-chains and the positional rule of percentages of *n*-plets

Similar phenomenological results were also received regarding percentages of *n*-plets in special subsequences of long nucleotide sequences of single-stranded DNA. These subsequences are termed «DNA epi-chains» [Petoukhov, 2019, 2020a-c]. Our initial results testify that the above-described equalities (2.1) of total sums of percentages of *n*-plets hold for these epi-chains as well. By definition, in a nucleotide sequence N_1 of any DNA strand with sequentially numbered nucleotides 1, 2, 3, 4, … (Fig. 8a), epi-chains of different orders *k* are such subsequences that contain only those nucleotides, whose numeration differ from each other by natural number $k = 1, 2, 3, 4, \ldots$For example, in any single-stranded DNA, one can consider its epi-chain of the second-order N_2, in which its nucleotide sequence numbers differ by $k = 2$: an epi-chain N_2 contains nucleotides with numerations 1, 3, 5, …(Fig. 8b). By analogy, an epi-chain of the third-order N_3 is connected with $k = 3$ and contains a subsequence of nucleotides with numerations 1, 4, 7, 10, … (Fig. 8c).

Each genomic DNA epi-chain of *k*th order (if $k = 2, 3, 4, \ldots$.) contains *k* times fewer nucleotides than the DNA strand and has its own arrangements of nucleobases A, T, C, and G. Each of these epi-chains contain different percentages of corresponding *n*-plets. But unexpectedly the total sums of percentages of *n*-plets in C-, G-, A-, and T-groupings practically coincide for each of the epi-chains and for the complete genomic DNA-text (at this stage of the research, the author studied the percentages of *n*-plets in epi-chains only in cases of epi-chains with relatively small orders *k*).

To illustrate this result, Fig. 9 shows the percent matrices for epi-chains of the second, third and fourth orders ($k = 2, 3, 4$) in the DNA of the human chromosome N☺1.

One can see from data in Fig. 9 that for considered epi-chains, different percentages matrices contain essential different percentages in corresponding separate cells. For example, %CG = 0.0103 in the complete single-stranded DNA; %CG = 0.0478 in the epi-chain of the 2nd order; %CG = 0.0464 in the epi-chain of the 3rd order; %CG = 0.0475 in the epi-chain of the 4th order. But in each of these four percentages matrices the equalities (2.1) are realized with a high level of accuracy:

$$[\%A,\%T,\%C,\%G] \approx [\sum \%AN, \sum \%TN, \sum \%CN, \sum \%GN] \approx [\sum \%NA, \sum \%NT, \sum \%NC, \sum \%NG] \approx [0.2910,\ 0.2918,\ 0.2085,\ 0.2087] \quad (3.1)$$

It illustrates that in the considered DNA epi-chains, the sums of the percentages in the tetra-groupings of *n*-plets in each of the epi-chains practically do not depend on the percent of its separate *n*-plets and

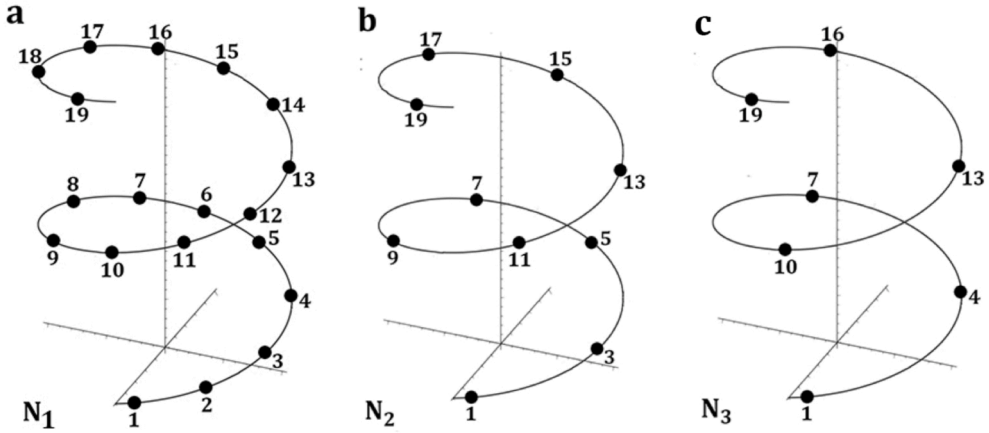

Fig. 8. Schematic representations of a single-stranded DNA and its initial epi-chains of numerated nucleotides, denoted by black circles. **a**, a sequence N_1 of numerated nucleotides of the DNA strand; **b**, an epi-chain of the second-order N_2 having nucleotides with numbers 1-3-5-7- …; **c**, an epi-chain of the third-order N_3 nucleotides numbers 1-4-7-10- ….

%CC	%CA	%AC	%AA
%CT	%CG	%AT	%AG
%TC	%TA	%GC	%GA
%TT	%TG	%GT	%GG

≈

0.05409	0.07274	0.05033	0.09504
0.07134	0.01031	0.07429	0.07137
0.06008	0.06312	0.04402	0.06008
0.09568	0.07286	0.05046	0.05419

Epi-chain of the 2nd order

0.0492	0.0524	0.0538	0.0985
0.0591	0.0478	0.0800	0.0588
0.0612	0.0788	0.0443	0.0613
0.0988	0.0530	0.0539	0.0492

Epi-chain of the 3rd order

0.0487	0.0562	0.0579	0.0912
0.0571	0.0464	0.0849	0.0570
0.0608	0.0826	0.0411	0.061
0.0918	0.0565	0.0578	0.0489

Epi-chain of the 4th order

0.0487	0.0558	0.0574	0.0940
0.0566	0.0475	0.0831	0.0565
0.0579	0.0835	0.0445	0.0579
0.0943	0.0561	0.0577	0.0487

Fig. 9. Matrices of percentages of 16 doublets in the single-stranded DNA of the human chromosome №1. Upper row: the percentages matrices of the genomic DNA ($k = 1$). Bottom row: percent matrices for the DNA epi-chains of the second, third, and fourth orders ($k = 2, 3, 4$).

coincide with corresponding sums in the genomic DNA-sequence. This can be considered as the separate Gestalt rule for genomic epi-chains (in addition to the above-presented Gestalt rule of percentages of *n*-plets in long multilayer DNA-sequences).

4. Regarding the development of Gestalt genetics based on analogies with Gestalt psychology

The described genomic phenomena of the relative independence of the sums of the percentages of *n*-plets (in the indicated tetra-groupings) from the values of individual percentages of summands in these sums resemble the phenomenon of perceiving a musical melody: a musical melody can be reproduced by different musical instruments and in different frequency ranges, that is, under significantly changing the sound frequency of each of its note elements, but despite these changes, the melody remains generally recognizable.

Many such phenomena of perception, in which there is relative independence of the integral form from its constituent individual components, are studied in Gestalt psychology. The described universal regularities in the preservation of total percentages in tetra-groupings of *n*-plets relatively regardless of the percentage of individual *n*-plets in genomic multilayer DNA-texts allow the author to develop Gestalt genetics. It seems natural to think that Gestalt genetics is interrelated with Gestalt psychology, which studies some genetically inherited properties of our brain regarding the perception of the environment.

The phenomena of perception of visual, auditory, and other images, studied by Gestalt psychology, reflect the fundamental inherited property of the psyche - to seek in a disparate whole. Thanks to the ability to think in Gestalts, you can understand the sentence, even if you change the order of the letters in each word and leave only the beginning and end in place. For example, you can easily understand the phrase, strongly "mutated" by local permutations: "*Aoccdrnig to a rscheearch at Cmabrigde Uinervtisy, it deosn't mttaer in waht oredr the ltteers in a wrod are*» (this "mutated" phrase example is taken from https://www.dictionary.com/e/typoglycemia/).

Gestalt genetics comes in contact with the teachings of the creator of analytical psychology C. Jung and his associate Nobel laureate in physics W. Pauli about the archetypes of the unconscious; in particular, they linked these archetypes with the Yin-Yang schemes of the ancient Chinese book "I-Ching" and its table of 64 hexagrams, which have deep structural analogies with the DNA alphabets (these analogies were noted by some reputable geneticists - G.S. Stent in 1969, and Nobel laureate F. Jacob in 1977; extended information on these analogies is in publications [Petoukhov, 1999, 2008; Petoukhov, He, 2010]).

In our opinion, the origins of the genetically inherited ability of the brain to work with Gestalt images should be sought in Gestalt genetics. In particular, Gestalt genetics is capable of providing new approaches to understanding the noise immunity of genetic information under mutations of DNA-texts.

In addition to the phenomena of Gestalt psychology, in living organisms, there are many genetically inherited physiological phenomena, in which the same whole pattern is realized in conditions of a wide variety of constituent elements and which can be attributed to Gestalt biology (this new name is proposed by the author as uniting genetically inherited Gestalt-like phenomena of different types). For example, Gestalt biology includes some genetically inherited phenomena of morphogenesis (laws of phyllotaxis, spiralization of biological structures at various levels and branches of biological evolution), as well as some functional phenomena (homeostasis at different stages of ontogenesis; the processing of sensory information from different sense organs according to main psychophysical law of Weber-Fechner; the biomechanical phenomenon, known as Bernstein's problem, that the general target task of body movement is performed exactly regardless of the inaccuracies of its constituent motor subtasks). In particular, Gestalt genetics includes an observation of the embryology classic K. Baer that chick embryos are vastly different, while the resulting adult organisms are remarkably similar. The author thinks that various genetically inherited phenomena of Gestalt biology are based on Gestalt genetics.

One should also recall that the molecular composition of a living body is constantly changing while maintaining the shape of the body. Our body's proteins are involved in continuous life-death cycles of their assembling and disassembling into amino acids. For example, the half-life of the hormone insulin is 6–9 min, etc. In other words, genetically inherited parts of our body are constantly dying and reborn. Taking into account such phenomena, the renowned physiologist A.G. Gurvich claimes: "*The main problem in biology is maintaining shape while constantly renewing the substrate*" [Gurvich, 1977]. In our opinion, the described phenomena of Gestalt genetics with its Gestalt rules of percentages in genomic multilayer DNA-texts are directly related to this fundamental problem of biology.

Gestalt genetics provides new approaches to understanding ontogeny. Using the terminology of Gestalt psychology and Gestalt therapy, the author suggests interpreting ontogenetic processes as the stepwise processes of "closing" certain genetic gestalts. One should note that all physiological systems are forced to bear the structural stamp of the genetic code since they should be genetically encoded for transmission to descendants and survival.

Gestalt genetics also concerns the phenomenon of the biomechanics of movements, described by the classic of biomechanics N.A. Bernstein: the general target task of the movement is performed exactly regardless of the inaccuracies of its constituent motor subtasks [Bernstein, 1967]. For example, when repeating an exact hit with a hammer on a nail, a person each time uses different trajectories, speeds, and accelerations of body parts with changes in both flexions in the joints and the activity of many muscles of each joint with many motoneurons of each muscle. This question of how the central nervous system is capable of adequately controlling the many degrees of freedom of the musculoskeletal system was first addressed by Bernstein and is now known usually as the

«Bernstein problem». The degrees of freedom problem in motor control states that there are multiple ways for humans or animals to perform a movement to achieve the same goal using redundant neurophysiological degrees of freedom. How the nervous system "chooses" a subset of these near-infinite degrees of freedom is an overarching difficulty in understanding motor control and motor learning. In other words, under normal circumstances, no simple one-to-one correspondence exists between a motor problem (or task) and a motor solution to the problem.

Gestalt psychology comes into contact with the well-known phenomena of perception constancy, reflecting the ability of the brain to stably recognize the shapes of objects and other structures of the external world under conditions of a significant change in the conditions of their presentation to the sense organs. For example, a change in the color illumination of a room does not prevent a person from recognizing the shapes and colors of objects in it, although other light frequencies from objects come to the retina. Similar properties one can be waiting for information phenomena in Gestalt genetics, which preseeds phenomena of Gestalt psychology.

According to Mendel's law of independent inheritance of traits, information from the level of DNA molecules dictates the macrostructures of living bodies through many independent channels, despite strong noises and interferences. For example, hair, eye, and skin colors are inherited independently of each other. Accordingly, each organism is a machine of multichannel noise-immune coding. Gestalt genetics, with its genomic Gestalt rules, can help in understanding this multi-channel noise immunity.

5. Gestalt genetics and quantum informatics

Many authors have long suggested the quantum-informational nature of living bodies. For example, R. Penrose, treating the body as a quantum computer, appeals to tubulin proteins that can switch from one state to another by analogy with triggers [Penrose, 1994]. One can show that the analysis of living bodies at a deeper - genomic - level leads to more convincing and fruitful pieces of evidence about organisms as quantum information entities (or as quantum-like entities whose modeling can use the formalisms of quantum informatics).

In quantum mechanics and quantum informatics, when analyzing the probabilities of events, the amplitudes of these probabilities, equal to the square root of their values, are traditionally considered. The genomic Gestalt rule for percentages of *n*-plets (in C-, G-, A-, and T-groupings) was revealed under analyzing the sums of the probabilities of *n*-plets from certain groupings, for example, all doublets starting with the letter C: $\Sigma\%CN = \%CC + \%CA + \%CT + \%CG$. Each of these probabilities %CC, %CA, %CT, and %CG has its own "amplitude" in the form of its square root.

It can be noted that the sum of the percentages (that is, the probabilities) of *n*-plets in each of these four groupings can be interpreted as the square of the length of a vector whose components are equal to the square roots of the probabilities of the corresponding *n*-plets. For example, the sum %CC+%CG+%CA+%CT is the square of the length of the 4-dimensional vector $V_{CN} = [\sqrt{\%CC}, \sqrt{\%CG}, \sqrt{\%CA}, \sqrt{\%CT}]$. Accordingly, the sum $\Sigma\%CNN$ is interpreted as the square of the length of an 8-dimensional vector $V_{CNN} = [\sqrt{\%CCC}, \sqrt{\%CCG}, \sqrt{\%CCA}, \sqrt{\%CCT}, \sqrt{\%CGC}, \sqrt{\%CGG}, \sqrt{\%CGA}, \sqrt{\%CGT}, \sqrt{\%CAC}, \sqrt{\%CAG}, \sqrt{\%CAA}, \sqrt{\%CAT}, \sqrt{\%CTC}, \sqrt{\%CTG}, \sqrt{\%CTA}, \sqrt{\%CTT}]$, and so on.

From this point of view, equalities (2.1) mean the constancy of the length of the corresponding 2^n-dimensional vectors, whose coordinates are the amplitudes of the probabilities of the corresponding *n*-plets. This metric approach allows for developing new methods of comparative vector analysis in genetics.

Let us compare all *m*-positional tetra-groupings - under different values *m* - in the case of real percentages of *n*-plets (similar to those shown in Figs. 3–5) and in the case of reference percentages (similar to those shown in Figs. 6 and 7). In all compared tetra-groupings, the corresponding 2^n-dimensional vectors of amplitudes of probabilities have different coordinate values. But the lengths of these vectors in all compared cases are equal to each other due to the Gestalt rule, which says about the equality of the sums of the percentages of *n*-plets in each grouping to the same value, that is, to the percentage of the corresponding nucleotide in the analyzed genomic DNA-text. Accordingly, these vectors of equal length can be transformed into each other by unitary transformations that do not change the length of the vectors and are either rotations or mirror reflections.

Thus, algebraic Gestalt genetics turns out to be connected with unitary operators, which are key for quantum informatics: all calculations in quantum computers and quantum search algorithms are based on unitary operators as quantum gates. Moreover, any unitary matrix can serve as a quantum gate. In quantum mechanics, the evolution of a closed quantum system is described by unitary transformations. Since this article deals with many layers of genomic DNA texts, it can be assumed that, in particular, a whole set of quantum search algorithms work in multilayer DNA texts, each of which is individually oriented to a particular text layer. The articles [Patel, 2001a-c] suppose that the genetic code is related to the quantum Grover's algorithm.

Along the way, one can note that the entire genetically inherited kinematic scheme of movements of our body with its parts is built on rotations in the joints and mirror reflections, that is, on unitary transformations that have physiological significance. Turtles and crocodiles, when hatched from an egg, immediately crawl to the water with coordinated movements using innate algorithms with the same unitary transformations of rotations and mirror reflections.

The materials of this new article supplement the author's previously published works on the topic of quantum biology and formalisms of quantum informatics in biology [Petoukhov, 2008, 2018, 2020a,c, 2021a,b; Petoukhov, He, 2010; Petoukhov et al., 2019]. The revealed connection of genetics with quantum informatics opens up the possibility of introducing rich formalisms of quantum mechanics and quantum informatics into algebraic biology for the mutual enrichment of these sciences and the inclusion of biology in the field of developed mathematical natural science. There are about 100 trillion cells in the human body, forming a holistic system. The formalisms of quantum informatics and Gestalt genetics can help in understanding such coherence phenomena.

The author draws special attention to unitary matrices in developing Gestalt genetics since he considers unitary rotation transformation as a basis for modeling a well-known morphogenetic Gestalt phenomenon of spiralization in biology that is, the existence of helical and spiral morphological configurations at different levels and branches of biological organization independently on their genetically inherited biomaterial content. Goethe even called spirals "symbols of life" because of multiple implementations of inherited spiral structures and processes in living bodies at various lines and levels of biological evolution. In the human body spiral and helical structures genetically inherited from one generation to another are presented in the muscles, heart, blood vessels, bones, nerves, an organ of hearing (the cochlea), cellular organization of the embryo (zygote), etc. The structure of tendons and ligaments consists of spirals and helices, which in turn are composed of collagen that has a triple helical structure. Spiral motions (nutation) are observed during the growth of roots and shoots, tendrils of plants are spirally wrapped, a tissue in the trunks of trees grows spirally, etc. Because of spiral bio-configurations, all the fluids in the body (blood, lymph, and urine) are spiral. The title of a book about bio-spirals – "Lines of Life" [Cook, 1914] - reflects their importance for living matter. Previous author's publications describe other examples of structural connections of molecular-genetic systems with unitary matrices as well [Petoukhov, 2008, 2018; Petoukhov, He, 2010].

In this article, the following new structural connection of unitary matrices with genomic DNA-texts is also noted. When considering a DNA double helix containing complementary nucleotide pairs C-G and A-T, the percentages of nucleotides C and G are exactly equal to each other

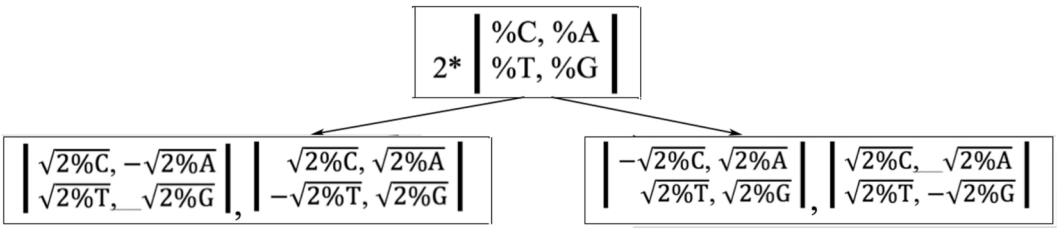

Fig. 10. The interrelation between the doubly stochastic matrix (upper row) of the percentage of nucleotides in the DNA double helix, where %C = %G and %A = %T, and four unitary matrices (bottom row). Two unitary matrices (at left) are transformations of rotation clockwise and counterclockwise and present complex number $z = \sqrt{2\%C} + i^*\sqrt{2\%A}$. The other two unitary matrices (at right) present transformations of mirror reflection.

(%C = %G), as are the percentages of nucleotides A and T (%A = %T). In this case, from the entire percentage sum %C + %A + %T + %G = 2(%C + %A) = 2(%C + %T) = 1.0, you have the following equalities: %A = 0.5 - %C and %T = 0.5 - %C. Correspondingly the percentage matrix 2[%C, %A; %T, %G] (known from Fig. 2) becomes a bisymmetric doubly stochastic matrix 2[%C, 0.5-%C; 0.5-%C, %C] where percentage sums in each row and each column are equal to 1.0. But doubly stochastic matrices are connected with unitary matrices as the following theorem claims [Prasolov, 1994]:

- if a square (n*n)-matrix M = $|m_{ij}|$ is unitary then a (n*n)-matrix B = $|b_{ij}|$, where $b_{ij} = |m_{ij}|^2$, is doubly stochastic.

In line with this theorem, Fig. 10 shows an interrelation between four unitary (2*2)-matrix and a doubly stochastic (2*2)-matrix of percentages of nucleotides in DNA double helices where %C = %G and %A = %T. Two of these four unitary matrices present transformations of mirror reflections. The other two unitary matrices are transformations of rotations of vectors clockwise and counterclockwise and simultaneously they are matrix presentations of complex number $z = \sqrt{2\%C} + i^*\sqrt{2\%A}$, where $i^2 = -1$ is the imaginary unit of complex numbers.

This interrelation of the structural features of DNA double helices, presented in the percentage matrix (upper row in Fig. 10), with unitary matrices of rotations clockwise and counterclockwise can be used for algebraic modeling of well-known Gestalt morphogenetic phenomena of left and right spiralizations realized in a great number of genetically inherited biological structures.

The described connection of matrices of nucleotide percentage in DNA double helices with complex numbers, which are presented by noted unitary matrices, is important for algebraic and quantum mechanical modeling of genetic phenomena since quantum mechanics is closely related to complex numbers.

Such interrelations of (2*2)-matrices, shown in Fig. 10, can be generalized for (2^n*2^n)-matrices produced by tensor powers of the initial unitary (2*2)-matrix and the doubly stochastic (2*2)-matrix (as it is known, tensor powers of unitary matrices generate new unitary matrices of increased orders). Tensor powers of (2*2)-matrices, which present complex numbers, produce (2^n*2^n)-matrices, which present algebraic extensions of complex numbers.

One should mention suppositions of many authors that formalisms of quantum informatics can be effectively used for deep understanding and modeling biological bodies (see, for example [Igamberdiev, 1993; Matsuno, 1999; Matsuno, Paton, 2000; Abbott et al., 2008; Fimmel, Petoukhov, 2020]:).

6. Gestalt percentage rules for sets of *n*-plets starting with certain k-plets (*k* < *n*)

Above, in section 2, m-positional tetra-groupings of n-plets were described, which are defined by a disposition of attributive nucleotides C, G, A, and T at a certain position m ($m \leq n$) inside these n-plets. Now in this section, other positional tetra-groupings of n-plets are described where a role of attributive elements is played by not separate nucleotides but separate doublets, or separate triplets, and so on. As it is revealed, these new tetra-groupings of n-plets obey Gestalt rules but have their own values for percentage sums in each of the groupings. Let us explain this by some examples using again percentages of n-plets in different layers of the DNA-text of the human chromosome №1 by analogy with section 2.

%AA ≈ Σ%AAN ≈ Σ%NAA ≈ Σ%AANN ≈ Σ%NAAN ≈ Σ%NNAA ≈ 0.095
%AT ≈ Σ%ATN ≈ Σ%NAT ≈ Σ%ATNN ≈ Σ%NATN ≈ Σ%NNAT ≈ 0.074
%AC ≈ Σ%ACN ≈ Σ%NAC ≈ Σ%ACNN ≈ Σ%NACN ≈ Σ%NNAC ≈ 0.050
%AG ≈ Σ%AGN ≈ Σ%NAG ≈ Σ%AGNN ≈ Σ%NAGN ≈ Σ%NNAG ≈ 0.071
%TA ≈ Σ%TAN ≈ Σ%NTA ≈ Σ%TANN ≈ Σ%NTAN ≈ Σ%NNTA ≈ 0.063
%TT ≈ Σ%TTN ≈ Σ%NTT ≈ Σ%TTNN ≈ Σ%NTTN ≈ Σ%NNTT ≈ 0.096
%TC ≈ Σ%TCN ≈ Σ%NTC ≈ Σ%TCNN ≈ Σ%NTCN ≈ Σ%NNTC ≈ 0.060
%TG ≈ Σ%TGN ≈ Σ%NTG ≈ Σ%TGNN ≈ Σ%NTGN ≈ Σ%NNTG ≈ 0.073
%CA ≈ Σ%CAN ≈ Σ%NCA ≈ Σ%CANN ≈ Σ%NCAN ≈ Σ%NNCA ≈ 0.073
%CT ≈ Σ%CTN ≈ Σ%NCT ≈ Σ%CTNN ≈ Σ%NCTN ≈ Σ%NNCT ≈ 0.071
%CC ≈ Σ%CCN ≈ Σ%NCC ≈ Σ%CCNN ≈ Σ%NCCN ≈ Σ%NNCC ≈ 0.054
%CG ≈ Σ%CGN ≈ Σ%NCG ≈ Σ%CGNN ≈ Σ%NCGN ≈ Σ%NNCG ≈ 0.010
%GA ≈ Σ%GAN ≈ Σ%NGA ≈ Σ%GANN ≈ Σ%NGAN ≈ Σ%NNGA ≈ 0.060
%GT ≈ Σ%GTN ≈ Σ%NGT ≈ Σ%GTNN ≈ Σ%NGTN ≈ Σ%NNGT ≈ 0.050
%GC ≈ Σ%GCN ≈ Σ%NGC ≈ Σ%GCNN ≈ Σ%NGCN ≈ Σ%NNGC ≈ 0.044
%GG ≈ Σ%GGN ≈ Σ%NGG ≈ Σ%GGNN ≈ Σ%NGGN ≈ Σ%NNGG ≈ 0.054

Fig. 11. Percentage sums are presented for mm-positional tetra-groupings related to 64 triplets and 256 tetraplets in the appropriate triplet- and tetraplet-layers of the DNA-text of the human chromosome №1. Each of these tetra-groupings is defined by one of 16 doublets as its attributive positional element disposed in the beginning, or in the middle, or at the end of the n-plets. Numerical values of percentage sums are calculated based on data about the percents of separate n-plets in Figs. 2–4.

For the beginning, consider the percentage contents of positional tetra-groupings, each of which is defined by one of 16 doublets as its attributive positional element. In a general case, one calls such tetra-groupings as *mm*-positional tetra-groupings or, in individual cases, as a CA-tetra-grouping, or a TA-tetra-grouping, etc. Below special denotations are used for percentage sums of *n*-plets, which are explained by the following examples (here N refers to any of nucleotides A, T, C, and G):

- Σ%TAN means a percentage sum of all 4 triplets, which start with the doublet TA in an analyzed DNA-text presented as a sequence of triplets;
- Σ%NTA means a percentage sum of all 4 triplets, which end with the doublet TA in an analyzed DNA-text presented as a sequence of triplets;
- Σ%TANN means a percentage sum of all 16 tetraplets, which start with the doublet TA in an analyzed DNA-text presented as a sequence of tetraplets;
- Σ%NTAN means a percentage sum of all 16 tetraplets, which have the doublet TA in their middle in a DNA-text presented as a sequence of tetraplets;
- Σ%NNTA means a percentage sum of all 16 tetraplets, which end with the doublet TA in an analyzed DNA-text presented as a sequence of tetraplets.

Let us calculate these percentage sums of *n*-plets for corresponding layers of the DNA-text of the human chromosome №1 using the data presented in Figs. 4–6 about percents of separate doublets, triplets, and tetraplets. This calculation gives equalities shown in Fig. 11.

Data in Fig. 11 show that percentage sums in each of these *mm*-tetra-groupings are equal to a percentage of that doublet, which plays the role of the attributive positional element for these tetra-groupings. The presented equality of these percentage sums in all tetra-groupings of each row of Fig. 11 occurs even though summands in them in each case are very different. Other genomes, analyzed by the author, have similar Gestalt properties for *mm*-positional tetra-groupings in the layers of their DNA-texts. This new genomic Gestalt phenomenon and the corresponding Gestalt rule for *mm*-positional tetra-groupings (Fig. 11) are analogs of the Gestalt phenomenon and the Gestalt rule described above for *m*-positional tetra-groupings in section 2 and the expression (2.1).

Now consider percentage contents of 64 mmm-positional tetra-groupings, each of which is defined by one of 64 triplets as its attributive positional element. The analogical denotations are used for these tetra-groupings:

- Σ%TAAN means a percentage sum of all 4 tetraplets, which start with the triplet TAA in an analyzed DNA-text presented as a sequence of tetraplets;
- Σ%NTAA means a percentage sum of all 4 tetraplets, which end with the triplet TAA.

%CCC ≈ Σ%CCCN ≈ Σ%NCCC ≈ 0.014	%ACC ≈ Σ%ACCN ≈ Σ%NACC ≈ 0.012
%CCG ≈ Σ%CCGN ≈ Σ%NCCG ≈ 0.003	%ACG ≈ Σ%ACGN ≈ Σ%NACG ≈ 0.003
%CGC ≈ Σ%CGCN ≈ Σ%NCGC ≈ 0.003	%AGC ≈ Σ%AGCN ≈ Σ%NAGC ≈ 0.014
%CGG ≈ Σ%CGGN ≈ Σ%NCGG ≈ 0.003	%AGG ≈ Σ%AGGN ≈ Σ%NAGG≈ 0.018
%GCC ≈ Σ%GCCN ≈ Σ%NGCC ≈ 0.013	%TCC ≈ Σ%TCCN ≈ Σ%NTCC ≈ 0.016
%GCG ≈ Σ%GCGN ≈ Σ%NGCG ≈ 0.003	%TCG ≈ Σ%TCGN ≈ Σ%NTCG ≈ 0.002
%GGC ≈ Σ%GGCN ≈ Σ%NGGC ≈ 0.013	%TGC ≈ Σ%TGCN ≈ Σ%NTGC ≈ 0.015
%GGG ≈ Σ%GGGN ≈ Σ%NGGG≈ 0.014	%TGG ≈ Σ%TGGN ≈ Σ%NTGG ≈ 0.019
%CCA ≈ Σ%CCAN ≈ Σ%NCCA ≈ 0.019	%ACA ≈ Σ%ACAN ≈ Σ%NACA ≈ 0.020
%CCT ≈ Σ%CCTN ≈ Σ%NCCT ≈ 0.019	%ACT ≈ Σ%ACTN ≈ Σ%NACT ≈ 0.016
%CGA ≈ Σ%CGAN ≈ Σ%NCGA ≈ 0.002	%AGA ≈ Σ%AGAN ≈ Σ%NAGA≈ 0.022
%CGT ≈ Σ%CGTN ≈ Σ%NCGT ≈ 0.003	%AGT ≈ Σ%AGTN ≈ Σ%NAGT ≈ 0.016
%GCA ≈ Σ%GCAN ≈ Σ%NGCA ≈ 0.015	%TCA ≈ Σ%TCAN ≈ Σ%NTCA ≈ 0.020
%GCT ≈ Σ%GCTN ≈ Σ%NGCT ≈ 0.014	%TCT ≈ Σ%TCTN ≈ Σ%NTCT ≈ 0.022
%GGA ≈ Σ%GGAN ≈ Σ%NGGA≈ 0.016	%TGA ≈ Σ%TGAN ≈ Σ%NTGA ≈ 0.019
%GGT ≈ Σ%GGTN ≈ Σ%NGGT ≈ 0.012	%TGT ≈ Σ%TGTN ≈ Σ%NTGT ≈ 0.020
%CAC ≈ Σ%CACN ≈ Σ%NCAC ≈ 0.015	%AAC ≈ Σ%AACN ≈ Σ%NAAC ≈ 0.014
%CAG ≈ Σ%CAGN ≈ Σ%NCAG ≈ 0.021	%AAG ≈ Σ%AAGN ≈ Σ%NAAG≈ 0.020
%CTC ≈ Σ%CTCN ≈ Σ%NCTC ≈ 0.018	%ATC ≈ Σ%ATCN ≈ Σ%NATC ≈ 0.013
%CTG ≈ Σ%CTGN ≈ Σ%NCTG ≈ 0.021	%ATG ≈ Σ%ATGN ≈ Σ%NATG ≈ 0.018
%GAC ≈ Σ%GACN ≈ Σ%NGAC ≈ 0.010	%TAC ≈ Σ%TACN ≈ Σ%NTAC ≈ 0.011
%GAG ≈ Σ%GAGN ≈ Σ%NGAG≈ 0.018	%TAG ≈ Σ%TAGN ≈ Σ%NTAG ≈ 0.013
%GTC ≈ Σ%GTCN ≈ Σ%NGTC ≈ 0.010	%TTC ≈ Σ%TTCN ≈ Σ%NTTC ≈ 0.020
%GTG ≈ Σ%GTGN ≈ Σ%NGTG ≈ 0.015	%TTG ≈ Σ%TTGN ≈ Σ%NTTG ≈ 0.019
%CAA ≈ Σ%CAAN ≈ Σ%NCAA ≈ 0.019	%AAA ≈ Σ%AAAN ≈ Σ%NAAA≈ 0.037
%CAT ≈ Σ%CATN ≈ Σ%NCAT ≈ 0.018	%AAT ≈ Σ%AATN ≈ Σ%NAAT ≈ 0.024
%CTA ≈ Σ%CTAN ≈ Σ%NCTA ≈ 0.013	%ATA ≈ Σ%ATAN ≈ Σ%NATA ≈ 0.019
%CTT ≈ Σ%CTTN ≈ Σ%NCTT ≈ 0.020	%ATT ≈ Σ%ATTN ≈ Σ%NATT ≈ 0.024
%GAA ≈ Σ%GAAN ≈ Σ%NGAA≈ 0.020	%TAA ≈ Σ%TAAN ≈ Σ%NTAA ≈ 0.020
%GAT ≈ Σ%GATN ≈ Σ%NGAT ≈ 0.013	%TAT ≈ Σ%TATN ≈ Σ%NTAT ≈ 0.019
%GTA ≈ Σ%GTAN ≈ Σ%NGTA ≈ 0.011	%TTA ≈ Σ%TTAN ≈ Σ%NTTA ≈ 0.020
%GTT ≈ Σ%GTTN ≈ Σ%NGTT ≈ 0.014	%TTT ≈ Σ%TTTN ≈ Σ%NTTT ≈ 0.037

Fig. 12. Percentage sums are presented for *mmm*-positional tetra-groupings related to 256 tetraplets in the tetraplet-layer of the DNA-text of the human chromosome №1. Each of these tetra-groupings is defined by one of 64 triplets as its attributive positional element disposed in the beginning or in the end of the tetraplets. Numerical values of percentage sums are calculated based on data about percents of separate triplets and tetraplets in Figs. 3–4.

Let us calculate these percentage sums of n-plets for corresponding layers of the DNA-text of the human chromosome N№1 using the data presented in Figs. 3 and 4 about the percents of separate triplets and tetraplets. This calculation gives equalities shown in Fig. 12.

Data in Fig. 12 show that percentage sums in each of these *mmm*-tetra-groupings are equal to a percent of the triplet, which plays the role of the attributive positional element for these tetra-groupings. The presented equality of these percentage sums in all tetra-groupings of each row of Fig. 12 occurs even though summands in them in each case are very different. Other genomes, analyzed by the author, have similar Gestalt properties for such *mmm*-positional tetra-groupings in the appropriate layers of their DNA-texts. This new genomic Gestalt phenomenon and the corresponding Gestalt rule for *mmm*-positional tetra-groupings (Fig. 12) are also analogs of the Gestalt phenomenon and the Gestalt rule described above for *m*-positional tetra-groupings in section 3 and the expression (2.1).

The author suggests that similar Gestalt phenomena exist also for layers of *n*-plets with $n = 5, 6, 7, \ldots$ of genomic DNA-texts of various types of organisms but he did not make corresponding calculations at this stage of research.

At the end of section 2, the concept of reference percentages of *n*-plets in the corresponding layers of genomic DNA texts was introduced based on the tensor family of matrices $[\%C, \%A; \%T, \%G]^{(n)}$. Figs. 6 and 7 showed these reference percentages of *n*-plets for the first layers of the DNA-text of the human chromosome N№1. It should be noted that for these reference percentages $[\%C, \%A; \%T, \%G]^{(n)}$ similar Gestalt rules hold for *mm*-positional and *mmm*-positional tetra-groupings in the appropriate layers of genomic DNA-texts. For example, for the human chromosome N№1 in line with its data in Figs. 6 and 7, you have the following equalities for a case when 16 doublets play the role of positional attributive elements:

$$\begin{aligned} &\%CCC + \%CCA + \%CCT + \%CCG \approx \%CC, \text{ because} \\ &0.00906382 + 0.01265035 + 0.01268315 + 0.00907448 \approx 0.0435 \\ &\%CAC + \%CAA + \%CAT + \%CAG \approx \%CA, \text{ because} \\ &0.01265035 + 0.01765605 + 0.01770183 + 0.01266522 \approx 0.0607 \end{aligned} \tag{6.1}$$

and so on for all doublets as attributive elements in *mm*-positional tetra-groupings.

7. The genomic rules and Frohlich's theory about collective quantum effects in biological systems

The presented results of the author's study indicate the presence of long-range relationships in the composition of genomic DNA-texts and also the quantum information essence of the genetic system and living organisms as a whole [Petoukhov, 2020a-c]. The author believes that these results are associated with the well-known vibration-resonance theory of G. Frohlich about long-range coherence in biological systems, that is, about collective quantum effects in biological systems [Frohlich, 1969, 1970, 1978, 1980, 1988; Frohlich, Kremer, 1983]. Let us clarify this belief.

Penrose uses Frohlich's theory to substantiate his ideas about the brain as a quantum computer and writes about it as follows: «*The distinguished physicist Herbert Frohlich suggested a possible role for collective quantum effects in biological systems. … Frohlich was led to propose, in 1968, that there should be vibrational effects within active cells, which would resonate with microwave electromagnetic radiation, at 10^{11} Hz, as a result of a biological quantum coherence phenomenon. Instead of needing a low temperature, the effects arise from the existence of a large energy of metabolic drive. There is now some respectable observational evidence, in many biological systems, for precisely the kind of effect that Frohlich had predicted in 1968 …. He argued that so long as the energy of metabolic drive is large enough, and the dielectric properties of the materials concerned are sufficiently extreme, then there is the possibility of large-scale quantum coherence similar to that which occurs in the phenomena of superconductivity and superfluidity - sometimes referred to as Bose-Einstein condensation – even at the relatively high temperatures that are present in biological systems … In a Bose-Einstein condensate, large numbers of particles participate collectively in a single quantum state. There is a wavefunction, for this state, of the kind that would be appropriate for a single particle - but now it applies all at once to the entire collection of particles that are participating in the state. … With a Bose-Einstein condensate, it is as though the entire system containing a large number of particles behaves as a whole very much as the quantum state of a single particle would, except that everything is scaled up appropriately. There is a coherence on a large scale, where many of the strange features of quantum wavefunctions hold at a macroscopic level*" [Penrose, 1994, c. 352 and 367].

F. Fröhlich (the son of Herbert Fröhlich) wrote the article "Genetic Code as Language" regarding quantum coherence states and long-range communication in genomes [Fröhlich F., 1988]. He notes: "*Beyond the chromosome, the genome as a whole must contain some sort of long-distance communication in order not to produce on one section antigens produced on other parts. So there is a complexity of entities which use genetic language in different ways. There might be said to be a logic of cells*" [Fröhlich F., 1988, p. 194]. He further writes that – by H. Fröhlich's hypothesis - long-range coherent vibrations will lead to resonance between a differentiated cell with its own characteristic vibrations and the chromosome such that the chromosome-particular region responding to this characteristic frequency will be activated or opened up so it can produce the appropriate proteins. Such a resonance could transport the embryological, already partially induced cells to their target and there they would be further fixed into producing the correct proteins for this organ through superimposed resonance. Frolich's theory allows an explanation of some features of cell division during mitosis by the hypothesis that corresponding chromosomes line up through resonance having the same frequency, finding each other by such long-range communication. Resonance oscillations draw like to like. More generally there might be some kind of coherent long-range interaction among the members of the genome creating the self-marking. If this broke down through subsequent mutations, auto-immune diseases would arise [Fröhlich F., 1988].

Fröhlich's synchronous large-scale collective oscillations imply intercellular microwave emissions which would constitute a non-chemical and non-thermal interaction between cells [Vasconcellos et al., 2012]. Some evidence of a non-thermal influence of coherent microwave radiation on the genome conformational state in *E. coli* has been reported [Hyland, 1998], which may indicate that chromosomal DNA could be the target of mm microwave irradiation within this system.

The genomic phenomenological rules described above can be considered as additional support of Frohlich's theory. The concept of multi-resonance genetics and its factological argumentation [Petoukhov, 2016] also has connections with Fröhlich's synchronous large-scale collective oscillations.

7.1. Some concluding remarks

The presented results of the study of the regularities in the percentage distribution of members of *n*-plets alphabets in genomic multi-level DNA-texts of various organisms are consistent with Jordan's claiming that life's missing laws are the rules of chance and probability of the quantum world [Jordan, 1932; McFadden, Al-Khalili, 2018]. The described author's results show the existence of previously unknown stochastic genetic regularities. These results were obtained using the described author's method of analysis and modeling of long DNA sequences as a set of many parallel texts based on n-plets alphabets.

Considering the views of Jordan and Schrödinger about the dictatorial role of the structured informatics of genetic molecules for the whole organism [McFadden, Al-Khalili, 2018], it is natural to think that the structural features of DNA informatics leave their marks on all genetically inherited biological systems and phenomena. This is consistent with the fact that all physiological systems must be structurally aligned with genetic coding to be transmitted in genetically

encoded forms to offspring. This is also consistent with the point of view that the main task of living organisms is to transfer genetic information along the chain of generations. The described algebraic-genetic results give pieces of evidence that the system of genetic coding is based on methods of probability coding.

The modern situation in the theoretic field of genetic informatics, where many millions of nucleotide sequences are described, can be characterized by the following citation: "*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. We have the program that runs the cellular machinery, but we know very little about how to read it*" [Fickett and Burks, 1989]. The important task of code biology is to provide progress in understanding languages of genetic messages and the described results on stochastic genomic rules connected with *n*-plets alphabets seem to be useful for this.

E.Schrodinger noted: "*from all, we have learned about the structure of living matter, we must be prepared to find it working in a manner that cannot be reduced to the ordinary laws of physics … because the construction is different from anything we have yet tested in the physical laboratory*» [Schrodinger, 1944]. For comparison, the enzymes in biological organisms work a million times more effectively than catalysts in the laboratory. What makes the enzyme in the body for 1 s, a catalyst in the laboratory can make only for 100 thousand years. The author believes that such ultra-efficiency of enzymes in biological bodies is defined not only by laws of physics, but also by quantum-logical relations in the genetic system, and therefore - in line with Schrodinger - this ultra-efficiency cannot be reduced to the ordinary laws of physics. As far as the author understands, the found Gestalt rules of "dictatorial" DNA-texts are not derived from the known laws of physics, and therefore refer to the special stochastic laws of the structuring of living things that Schrödinger spoke about.

Regarding the metric features of biological phenomena (in some connection with mentioned unitary matrices and algebras of hypercomplex numbers), it can be noted that thoughts about metric spaces are spread by some authors even to mathematics as a high form of intellectual activity. For example, the book [Hofstadter, 1980, p. 612] notes that a mathematician feels that in mathematics there is a certain metric that unites ideas - that all mathematics is a network of results that are interconnected by a huge number of connections; had we been able to introduce this highly developed sense of mathematical closeness - the mental metric of a mathematician - into the program, we could create a primitive artificial mathematician.

By these statements of Hofstadter, the book [Nalimov, 2015, p. 115] emphasizes: "*In other words, artificial intelligence could be brought closer to mathematical thinking, if it were possible to realize the metric properties of the human thinking space … We are ready to go further and say that consciousness itself is geometrically structured: existentially, a person is geometric … In our minds, when constructing texts through which we perceive the World, something very similar to what happens in morphogenesis happens. We are ready to see in the depths of consciousness the same geometric images that are revealed in morphogenesis*".

In general, the presented results about the stochastic rules of genetic informatics give pieces of evidence in favor of the effectivity of a model approach to living organisms as quantum-informational algebraic-harmonic essences on modular principles. The approach and results are useful for understanding the important role of a wide set of interrelated genetic alphabets and further developing science on code biology [Barbieri, 2015].

Acknowledgments

Some results of this paper have been possible due to long-term cooperation between Russian and Hungarian Academies of Sciences on the theme "Non-linear models and symmetrologic analysis in biomechanics, bioinformatics, and the theory of self-organizing systems", where the author was a scientific chief from the Russian Academy of Sciences. The author is grateful to G. Darvas, E. Fimmel, A.A. Koblyakov, M. He, Z.B. Hu, Yu.I. Manin, I.V. Stepanyan, V.I. Svirin, and G.K. Tolokonnikov for their collaboration.

Appendix I. On the tensor product of matrices

The tensor (or Kronecker) product of matrices, denoted by ⊗, is widely used in mathematics, theoretical physics, informatics, control theory, etc. In a general case, if K_1 is an (m*n)-matrix and K_2 is a (p*q)-matrix, then the tensor product $K_1 \otimes K_2$ is the (mp*nq)-block matrix:

$$K_1 \otimes K_2 = \begin{vmatrix} a_{11}K_2 & \ldots. & a_{1n}K_2 \\ . & . & . \\ . & . & . \\ . & . & . \\ a_{m1}K_2 & \ldots. & a_{mn}K_2 \end{vmatrix}$$

Fig. A1. The definition of the tensor product of two matrices K_1 and K_2.

The tensor product is the crucial operation to understanding the quantum mechanics of multiparticle systems and is one of basic instruments in quantum informatics. The following quotation speaks about the tensor product: «*This construction is crucial to understanding the quantum mechanics of multiparticle systems*» [Nielsen, Chuang, 2010, p. 71] since in line with the postulate of quantum mechanics: the state space of a composite system is the tensor product of the state spaces of its components.

This operation has the following important property for square matrices: if a (*n*n*)-matrix K_1 has eigenvalues s_j (j = 1, …, *n*) and another (*m*m*)-matrix K_2 has eigenvalues g_i (i = 1, …, *m*), then the tensor product of these matrices $K_1 \otimes K_2$ has eigenvalues $s_j * g_i$ [Bellman, 1960]. Figuratively speaking, tensor products of matrices are endowed with the property of "inheritance" of the eigenvalues of these matrices in this form. This property is used in the concept of multi-resonance genetics [Petoukhov, 2016] and in modeling inherited physiological structures in Mendelian genetics [Petoukhov, 2011; 2021a].

Let us return to the genetic tables of the DNA alphabets of the *n*-plets in Fig. 1, which were built on the basis of binary-oppositional indicators in the alphabet of 4 nucleotides. Using the definition of the tensor product (Fig. A1), one can check that matrices, which are generated by corresponding tensor powers of the matrix [C, A; T, G] of 4 nucleotides (Fig. A2), are identical to the tables of 16 doublets, 64 triplets, 256 tetraplets from Fig. 1.

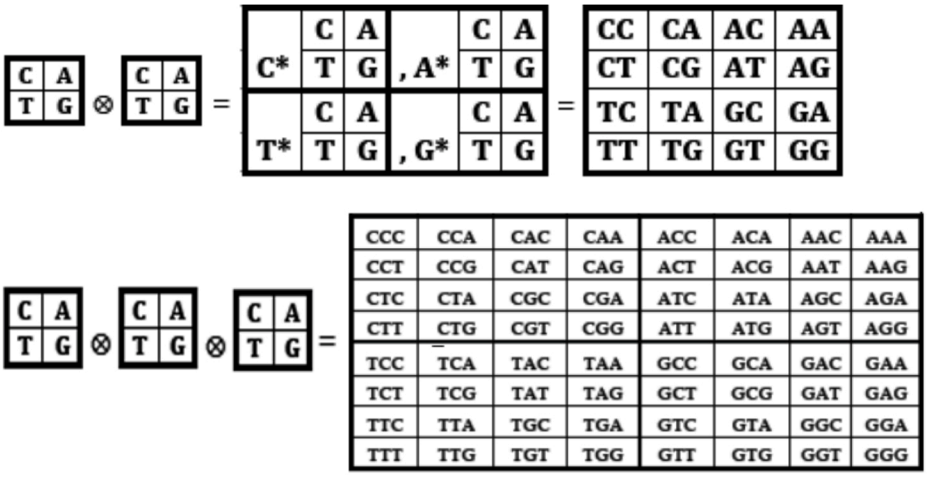

Fig. A2. Examples of generating DNA matrices of alphabets of *n*-plets by exponentiation of the matrix of four nucleotides [C, A; T, G] to corresponding tensor powers (compare with the alphabetical tables at Fig. 1).

References

Abbott, D., Davies, P.C.W., Pati, A.K., 2008. Quantum Aspects of Life (Foreword by Sir Roger Penrose), First. Imperial College Press, London, pp. 1–442.

Albrecht-Buehler, G., 2006. Asymptotically increasing compliance of genomes with Chargaff's second parity rules through inversions and inverted transpositions. Proc Natl Acad Sci U S A 103 (47), 17828–17833.

Barbieri, M., 2015. Code Biology. A New Science of Life. Springer, ISBN 978-3-319-14534-1. https://doi.org/10.1007/978-3-319-14535-8.

Bellman, R., 1960. Introduction to Matrix Analysis. Mcgraw-Hill Book Comp, N-Y.

Bernstein, N.A., 1967. The Co-ordination and Regulation of Movements. Pergamon Press, Oxford.

Capra, F., 2010. The Tao of Physics: an Exploration of the Parallels Between Modern Physics And Easterm Mysticism. Shambala, Boston USA.

Chargaff, E., 1971. Preface to a Grammar of Biology: a hundred years of nucleic acid research. Science 172, 637–642.

Cook, T.A., 1914. The Curves of Life. Constable and Co, London.

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

Fimmel, E., Petoukhov, S.V., 2020. Development of models of quantum biology based on the tensor product of matrices. In: Hu, Z., Petoukhov, S., He, M. (Eds.), Advances in Intelligent Systems and Computing, vol. 1126. Springer, Cham, pp. 126–135. https://doi.org/10.1007/978-3-030-39162-1_12.

Fröhlich, F., 1988. Genetic code as Language. In: Fröhlich, H. (Ed.), Biological Coherence And Response To External Stimuli. Springer, ISBN 978-0-387-18739-6, pp. 192–204.

Fröhlich, H., 1969. Quantum mechanical concepts in biology. In: Marois, M. (Ed.), From Theoretical Physics To Biology. North-Holland, Amsterdam, the Netherlands, pp. 13–22.

Fröhlich, H., 1970. Long range coherence and the action of enzymes. Nature 228, 1093. https://doi.org/10.1038/2281093a0, 1970.

Fröhlich, H., 1978. Coherent electric vibrations in biological systems and cancer problem. IEEE Trans. Microw. Theor. Tech. 26, 613–617.

Fröhlich, H., 1980. The biological effects of microwaves and related questions. Adv. Electron. Electron. Phys. 53, 85–152 (New York, USA: Academic Press).

Fröhlich, H., 1988. Introduction. Theoretical physics and biology. In: Herbert Fröhlich Editor, *Biological Coherence And Response To External Stimuli*. Springer, ISBN 978-0-387-18739-6, pp. 3–24, 1988.

Fröhlich, H., Kremer, F., 1983. Coherent Excitations in Biological Systems. Springer-Verlag, ISBN 978-3-642-69186-7.

Gurvich, A.G., 1977. Selected Works. Medicine, Moscow.

Hofstadter, D.R., 1980. Gödel, Escher, Bach. An Eternal Golden Braid. Vintage Books, N. Y., p. 777, 1980.

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. Prog. Biophys. Mol. Biol. 131, 354–368.

Hyland, G.J., 1998. Coherent GHz and THz excitations in active biosystems, and their implications. In: The Future of Medical Diagnostics?» Proceeding of Matra Marconi UK, Directorate of Science, Internal Report, Portsmouth, UK, 25 June 1998, pp. 14–27.

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

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

Karlin, S., Ost, F., Blaisdell, B.E., 1989. Patterns in DNA and amino acid sequences and their statistical significance. In: Waterman, M.S. (Ed.), Mathematical Methods for DNA Sequences. CRC Press, Florida.

Matsuno, K., 1999. Cell motility as entangled quantum coherence. Biosystems 51, 15–19.

Matsuno, K., Paton, R.C., 2000. Is there a biology of quantum information? Biosystems 55, 39–46, 2000.

McFadden, J., Al-Khalili, J., 12 December 2018. The origins of quantum biology. Proc. R. Soc. A 474 (2220), 1–13. https://doi.org/10.1098/rspa.2018.0674, 2018.

Nalimov, V.V., 2015. *I Am Scattering Thoughts* (In Russian: *Razbrasyvaiu Mysli*). Center for Humanitarian Initiatives, Moscow, ISBN 978- 5-98712-521-2, 2015, in Russian.

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

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

Patel, A., 2001b. Testing quantum dynamics in genetic information processing. J. Genet. 80 (1), 39–43.

Patel, A., 2001c. Why genetic information processing could have a quantum basis. J. Biosci. 26 (2), 145–151.

Penrose, R., 1994. Shadows of the Mind: A Search For the Missing Science of Consciousness. Oxford University Press, USA, p. 480.

Petoukhov, S.V., 1999. Genetic code and the ancient Chinese «book of Changes». Symmetry:Cult. Sci. 10 (№3–4), 211–226.

Petoukhov, S.V., 2008. Matrix Genetics, Algebrases of Genetic Code, Noise Immunity. RCD, Moscow, ISBN 978-5-93972-643-6, p. 316 (in Russian).

Petoukhov, S.V., 2011. Matrix genetics and algebraic properties of the multi-level system of genetic alphabets. NeuroQuantology 9 (4), 60–81.

Petoukhov, S.V., 2016. The system-resonance approach in modeling genetic structures. Biosystems 139, 1–11.

Petoukhov, S.V., 2018. The genetic coding system and unitary matrices. Preprints 2018040131. https://doi.org/10.20944/preprints201804.0131.v2, 2018.

Petoukhov, S.V., 2019. Nucleotide epi-chains and new nucleotide probability rules in long DNA sequences, 2019040011 Preprints 17, 10.20944/preprints201904.0011.v1. https://www.preprints.org/manuscript/201904.0011/v1.

Petoukhov, S.V., 2020a. Hyperbolic rules of the cooperative organization of eukaryotic and prokaryotic genomes. Biosystems 198, 104273.

Petoukhov, S.V., 2020b. Hyperbolic rules of the oligomer cooperative organization of eukaryotic and prokaryotic genomes. Preprints 2020, 2020050471. https://doi.org/10.20944/preprints202005.0471.v2. https://www.preprints.org/manuscript/202005.0471/v2.

Petoukhov, S.V., 2020c. The Rules of Long DNA-Sequences and Tetra-Groups of Oligonucleotides arXiv:1709.04943v6 , 6th version from. (Accessed 22 May 2020).

Petoukhov, S.V., 2021a. Modeling inherited physiological structures based on hyperbolic numbers. Biosystems 199, 104285. https://doi.org/10.1016/j.biosystems.2020.104285. ISSN 0303-2647.

Petoukhov, S.V., 2021b. Algebraic rules for the percentage composition of oligomers in genomes, 2021010360 Preprints 2021, 58. https://doi.org/10.20944/

preprints202101.0360.v2. from 1 March. https://www.preprints.org/manuscript/202101.0360/v2.
Petoukhov, S.V., He, M., 2010. Symmetrical Analysis Techniques for Genetic Systems and Bioinformatics: Advanced Patterns and Applications. IGI Global, Hershey, USA.
Petoukhov, S.V., Petukhova, E.S., Svirin, V.I., 2019. Symmetries of DNA alphabets and quantum informational formalisms. Symmetry: Cult. Sci. 30 (2), 161–179. https://doi.org/10.26830/symmetry_2019_2_161.
Prabhu, V.V., 1993. Symmetry observation in long nucleotide sequences. Nucleic Acids Res 21, 2797–2800.
Prasolov, V.V., 1994. Problems and Theorems in Linear Algebra (Translations of Mathematical Monographs, vol. 134. American Mathematical Society, p. 225.
Schrödinger, E., 1944. What Is Life? Cambridge Press.
Shchutskii, Y.K., 1997. The Chinese Classical "I Ching". Moscow, Vostochnaya literatura (in Russian).
Stewart, I., 1999. Life's Other Secret: the New Mathematics of the Living World. Penguin, New- York.
Vasconcellos, A.R., Vannucchi, F.S., Mascarenhas, S., Luzzi, R., 2012. Frohlich condensate: emergence of synergetic dissipative structures in information processing biological and condensed matter systems. Information 3 (4), 601–620. https://doi.org/10.3390/info3040601.