

Nucleotide epi-chains and new nucleotide probability rules in long DNA sequences

Sergey V. Petoukhov¹

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

Comment: Some elements of this article were presented by the author in his keynote speeches at the following conferences: the International Belgrade Bioinformatics Conference 2018 (Belgrade, Serbia, 18-22 June, 2018, <http://belbi.bg.ac.rs/>); the Second International Conference “Artificial Intelligence, Medical Engineering, Education” (Moscow, Russia, 6-8 October 2018, <http://www.ruscncf.org/aimee2018/index.html>); the First Congress of the International Society of Natural Medicine (Samorin, Slovak Republic, 12-14 October 2018, <https://www.simn.info/congress-of-natural-medicine-2018/>); the Second International Conference on Computer Science, Engineering and Education Applications (Kiev, Ukraine, 27-28 January 2019, <http://www.uacnconf.org/iccseea2019/index.html>).

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. The article presents author's results of studying probabilities of nucleotides on so-called epi-chains of long DNA sequences of various eukaryotic and prokaryotic genomes. DNA epi-chains are algorithmically constructed subsequences of DNA nucleotide sequences. According to the algorithm of construction of any epi-chain of the order n , the epi-chain is such nucleotide subsequence, in which the numerations of adjacent nucleotides differ by n ($n = 2, 3, 4, \dots$). Correspondingly each epi-chain of order n contains n times less nucleotides than the original DNA sequence. The presented results unexpectedly show that nucleotide probabilities on such DNA epi-chains of different orders are practically identical to nucleotide probabilities in the original long DNA sequence. These data allow considering DNA as a regular rich set of epi-chains, which can play a certain role in genetic and epigenetic phenomena as the author believes. Appropriate rules of nucleotide probabilities on epi-chains of long DNA sequences are formulated for further their tests on a wider set of biological genomes. These phenomenological data and their possible biological meaning are discussed.

Keywords: DNA sequence, probability, DNA epi-chains, epigenetics, quantum biology.

Content

1. Introduction
 2. Definitions of notions of code-chains and epi-chains of nucleotides in DNA
 3. Nucleotide probabilities in epi-chains of complete sets of chromosomes
 4. Nucleotide probabilities in epi-chains of prokaryotic genomes
 5. Regarding probabilities of triplets in DNA epi-chains of different orders
 6. Regarding a supposed genetic role of epi-chains
 7. Some concluding remarks
- Acknowledgments
References

1. Introduction

One of creators of quantum mechanics P. Jordan, credited with the authoring the first work on quantum biology, postulated the following. The mechanisms of living organisms are associated with what he referred to as his ‘amplifier theory’, based on Bohr's notion of the ‘irreversible act of amplification’, required to bring the fuzzy quantum reality into sharp focus by ‘observing’ it [Jordan, 1932]. Jordan claimed that *«life's missing laws were the rules of chance and probability (the indeterminism) of the quantum world that were*

somehow scaled up inside living organisms» [McFadden, Al-Khalili, 2018]. It is these laws of chance and probability, postulated by Jordan, that we are looking for in our studies of the probabilistic characteristics of long DNA sequences of hydrogen bonds and nitrogenous bases [Petoukhov, 2018a,b, 2019a,b; Petoukhov, Petukhova, Svirin, 2018; Petoukhov, Svirin, 2012, 2018; Darvas, 2018]. Long DNA sequences are a very important and convenient for such computerized research of possible biological rules of probabilities, which in Jordan's day was impossible. The main idea of all quantum mechanics is that everything in the world of appropriate objects is described only as probabilistic. Correspondingly the author has undertaken study of probabilities of nucleotides in long DNA of many species on the basis of a new author's approach.

Regarding probabilities of nucleotides in DNA, the so-called Chargaff's rules have been known since the middle of the last century [Chargaff, Lipshitz, Green, 1952; Chargaff, 1971]. In DNA molecules, genetic information is written in very long sequential texts using only 4 letters: adenine A, cytosine C, guanine G, thymine T. For example, the human genome consists of several billion such letters. The first Chargaff's rule holds that in double-stranded DNA molecule the amount of guanine (G) is equal to the amount of cytosine (C) and the amount of adenine (A) is equal to the amount of thymine (T). This rule was theoretically confirmed by the double-helix model of the DNA [Watson, Crick, 1953]. The second Chargaff's rule states that the same parity is approximately valid ($\%C \approx \%G$ and $\%A \approx \%T$) for each of the two long DNA strands alone (the term "long" usually refers to DNA sequences containing ≥ 100000 nucleotides). According to [Albrecht-Buehler, 2006], this rule applies to the eukaryotic chromosomes, the bacterial chromosomes, the double stranded DNA viral genomes, and the archaeal chromosomes provided they are long enough. Some our works were also devoted to study probabilities of nucleotides and hydrogen bonds in long DNA sequences [Petoukhov, 2018a,b, 2019; Petoukhov, Petukhova, Svirin, 2018; Petoukhov, Svirin, 2018; Darvas, 2018].

This article presents, first of all, the author's results of calculation of probabilities (or frequencies) of nucleotides located in algorithmically sparsed sequences inside long DNA of many eukaryotic and prokaryotic organisms. These numeric results testify in favor of existence of some general rules for long DNA nucleotide sequences and moreover they lead to a new important notion of "DNA epi-chains" defined below.

2. Definitions of notions of code-chains and epi-chains of nucleotides in DNA

DNA double helix is a long discrete polyatomic construction, whose parts are connected into the whole stabilized structure. In this construction, each of two complementary strands contains a discrete chain of nucleotides, which carry genetic information and are located along helix-like trajectory (Fig. 1, left).

The article considers two types of nucleotide chains, which are conditionally termed as "code-chains" and "epi-chains". The attention paid to these two types of chains in DNA is caused by the author's results on the approximate equality of the probabilities under investigation on these different chains.

A chain of nucleotides, which continuously follow each other on one strand of DNA, will be called a code-chain (the concept of a DNA strand is broader than the concept of its nucleotide code-chain, since the DNA strand contains not only nucleotides, but also the connections between them). The DNA double helix consists of two strands, and therefore contains two code-chains (one for each strand). It is precisely the nucleotide code-chains are used in the mentioned second Chargaff's rule, as well as the data on the nucleotide composition of the DNA of the set of organisms in the Genbank. The consecutive nucleotides in any code-chain can be sequentially numerated with the natural numbers 1, 2, 3, 4, 5, ..., starting with any randomly chosen nucleotide (Fig. 1). Below we will deal with nucleotides, each of which is conditionally assigned a sequence number along the nucleotide sequence in a considered code-chain.

In contrast to the traditional attention of researchers to DNA code-chains, this article focuses on epi-chains, the concept of which is new for molecular genetics. By definition, DNA nucleotide epi-chains are those nucleotide chains of a double or single DNA helix, which contain not all, but only algorithmic selected nucleotides, whose sequence numbers increase on a fixed number "n"; this numerical difference "n" can be equal to 2, 3, 4, ... and it defines the "order of epi-chains". For example, epi-chains of the second order are two sequences $N_{2/1}$ and $N_{2/2}$ whose nucleotides have their sequence numbers differing by 2, that is, they have odd numerations 1, 3, 5, 7, ... in an epi-chain $N_{2/1}$ and they have even numerations 2, 4, 6, 8, ... in an epi-chain $N_{2/2}$ as shown in Fig. 1 (in the middle). By analogy, epi-chains of the third order are those three nucleotide chains $N_{3/1}$, $N_{3/2}$ and $N_{3/3}$, whose nucleotides have sequence numbers differing by 3, that is, they have the numerations 1, 4, 7, 10, ... or 2, 5, 8, 11, ... or 3, 6, 9, 12, ... , respectively (Fig.1, right).

In the general case, nucleotide epi-chains of n -th order are those « n » chains, whose nucleotides have sequence numbers differing by « n ». The nucleotide density (or quantity) on a epi-chain of n -th order is reduced by " n " times in comparison with the nucleotide density on a corresponding DNA code-chain. In other words, DNA epi-chains are sparsed code-chains. The term "epi-chain" uses the Ancient Greek prefix epi- (ἐπι- "over, outside of, around"), which implies features that are "on top of" or "in addition to" the traditional DNA double helix. In DNA double helix, each nucleotide belongs to one of its code-chains and to many its epi-chains.

The symbol N in the designation of code-chains and epi-chains is taken according to the first letter of the word "nucleotides". In the symbolic designation of epi-chain $N_{n/m}$, the numerator " n " in the chain index indicates the order of the epi-chain, and the denominator " m " is the number of the nucleotide in the code-chain, from which the epi-chain begins. For example, the symbol $N_{3/2}$ means an epi-chain of the third order with the first nucleotide having the number 2 in nucleotide numeration in the code-chain.

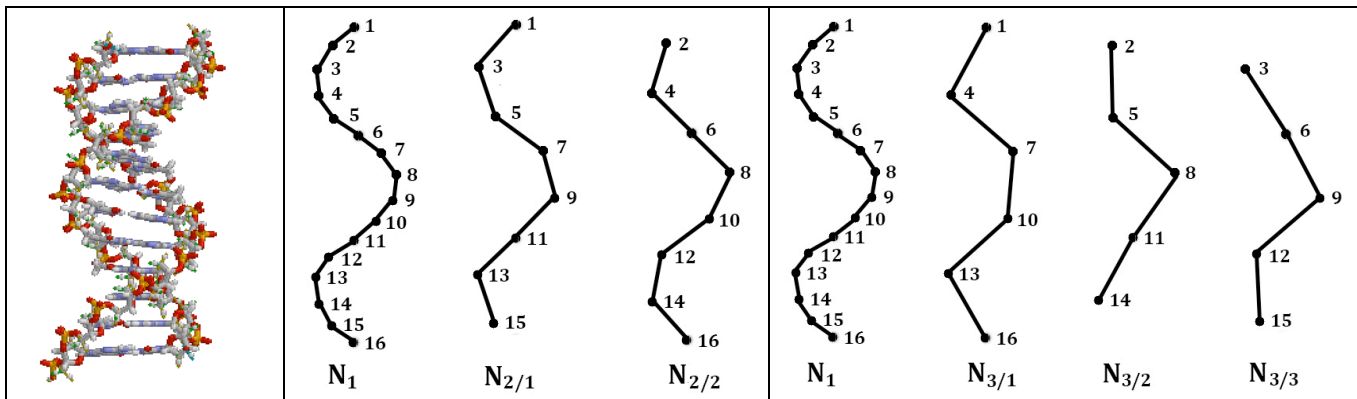


Fig. 1. Left: DNA double helix (from https://ru.wikipedia.org/wiki/Дезоксирибонуклеиновая_кислота). In the middle: N_1 represents a chain of numerated nucleotides of single-stranded DNA termed by the author as a DNA code-chain; $N_{2/1}$ and $N_{2/2}$ represent two epi-chains of the second order in this single-stranded DNA. Right: $N_{3/1}$, $N_{3/2}$ and $N_{3/3}$ represent three epi-chains of the third order regarding the same code-chain N_1 (explanations in the text).

This article presents data on the author's analysis of probabilities of nucleotides A, T, C, and G in epi-chains of long single-stranded DNA of genomes of a number of organisms.

3. Nucleotide probabilities in epi-chains of complete sets of chromosomes

This paragraph presents the author's study results that probabilities of nucleotides in the code-chains and epi-chains of the considered orders in long DNA sequences unexpectedly turn out to be almost identical independently on strong differences in content of these nucleotide chains. These probabilities (or frequencies) of nucleotides A, T, C and G in the studied DNA sequence will be denoted by the symbols $P(A)$, $P(T)$, $P(C)$ and $P(G)$, respectively.

Let us describe, first of all, data about the human genome with its 22 autosomes and 2 sex chromosomes. These 24 nuclear chromosomes contain long DNA molecules, the lengths of texts in which lie in the range from 50 to 250 million letters approximately. These chromosomes differ greatly in their molecular dimensions, their sequences of letters, kinds and quantities of genes in them, cytogenetic bands (which shows biochemical specificity of different parts of chromosomes), etc. (Fig. 2).

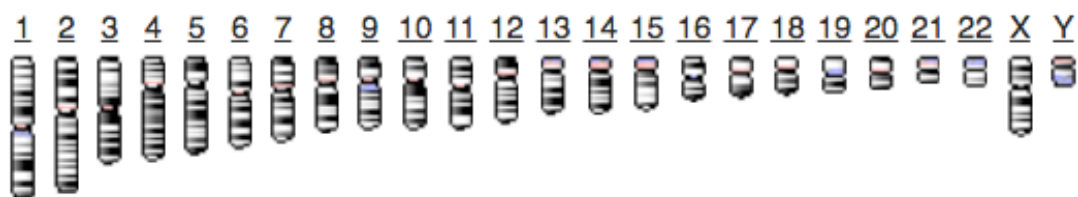


Fig. 2. Human chromosomes (<https://www.ncbi.nlm.nih.gov/genome/51>)

Taking into account these great differences among human chromosomes, it was very unexpected for the author to reveal that 24 human chromosomes are similar each to other from the standpoint of nucleotide probabilities in their DNA epi-chains (Table 1).

Table 1. Probabilities P(A), P(T), P(C) and P(G) of nucleotides A, T, C and G in epi-chains of initial orders in all human nuclear chromosomes. Data relating epi-chains of the second order ($N_{2/1}$, $N_{2/2}$) and the third order ($N_{3/1}$, $N_{3/2}$, $N_{3/3}$) are presented jointly with data on code-chains N_1 . Two left columns show ordinary symbols of chromosomes and their length. Initial data relating to these chromosomes were accessed from <https://www.ncbi.nlm.nih.gov/genome/?term=Homo+sapiens+genome> (the column RefSeq).

Chr	DNA length (bp)		N_1	$N_{2/1}$	$N_{2/2}$	$N_{3/1}$	$N_{3/2}$	$N_{3/3}$
1	248956422	P(A)	0.2910	0.2910	0.2910	0.2910	0.2910	0.2910
		P(T)	0.2918	0.2917	0.2918	0.2917	0.2917	0.2918
		P(C)	0.2085	0.2085	0.2085	0.2084	0.2085	0.2085
		P(G)	0.2087	0.2088	0.2087	0.2088	0.2088	0.2087
2	242193529	P(A)	0.2984	0.2985	0.2984	0.2983	0.2985	0.2985
		P(T)	0.2993	0.2992	0.2993	0.2993	0.2992	0.2993
		P(C)	0.2009	0.2009	0.2009	0.2009	0.2009	0.2008
		P(G)	0.2014	0.2014	0.2014	0.2014	0.2014	0.2014
3	198295559	P(A)	0.3013	0.3013	0.3013	0.3013	0.3013	0.3014
		P(T)	0.3020	0.3020	0.3020	0.3022	0.3020	0.3019
		P(C)	0.1980	0.1981	0.1980	0.1979	0.1981	0.1981
		P(G)	0.1986	0.1986	0.1986	0.1986	0.1986	0.1986
4	190214555	P(A)	0.3086	0.3086	0.3086	0.3087	0.3086	0.3086
		P(T)	0.3089	0.3090	0.3089	0.3089	0.3090	0.3090
		P(C)	0.1910	0.1909	0.1910	0.1909	0.1910	0.1910
		P(G)	0.1915	0.1915	0.1914	0.1915	0.1914	0.1914
5	181538259	P(A)	0.3018	0.3018	0.3018	0.3018	0.3017	0.3018
		P(T)	0.3032	0.3032	0.3032	0.3031	0.3032	0.3032
		P(C)	0.1971	0.1971	0.1971	0.1971	0.1972	0.1971
		P(G)	0.1979	0.1979	0.1980	0.1980	0.1979	0.1979
6	171718000	P(A)	0.3021	0.3021	0.3021	0.3021	0.3021	0.3022
		P(T)	0.3020	0.3020	0.3021	0.3022	0.3020	0.302
		P(C)	0.1979	0.1979	0.1979	0.1979	0.1980	0.1979
		P(G)	0.1979	0.1979	0.1979	0.1979	0.1979	0.1979
7	159345973	P(A)	0.2960	0.2960	0.2960	0.2960	0.2960	0.296
		P(T)	0.2970	0.2970	0.2970	0.2971	0.2970	0.297
		P(C)	0.2033	0.2033	0.2033	0.2032	0.2034	0.2033
		P(G)	0.2037	0.2037	0.2037	0.2037	0.2037	0.2037
8	146399655	P(A)	0.2994	0.2994	0.2994	0.2994	0.2994	0.2994
		P(T)	0.2990	0.2990	0.2989	0.2990	0.2990	0.2990
		P(C)	0.2008	0.2008	0.2008	0.2008	0.2008	0.2008
		P(G)	0.2008	0.2008	0.2009	0.2008	0.2008	0.2008
9	111583154	P(A)	0.2928	0.2928	0.2928	0.2927	0.2928	0.2929
		P(T)	0.2926	0.2927	0.2925	0.2926	0.2926	0.2927
		P(C)	0.2074	0.2073	0.2074	0.2074	0.2074	0.2073
		P(G)	0.2072	0.2072	0.2073	0.2073	0.2072	0.2072
10	133797422	P(A)	0.2917	0.2918	0.2916	0.2917	0.2917	0.2918
		P(T)	0.2929	0.2928	0.2929	0.2929	0.2928	0.2929
		P(C)	0.2074	0.2073	0.2075	0.2074	0.2075	0.2073

		P(G)	0.2080	0.2080	0.2080	0.2080	0.2080	0.2080
11	135086622	P(A)	0.2920	0.2920	0.2921	0.2920	0.2921	0.2920
		P(T)	0.2926	0.2926	0.2926	0.2925	0.2926	0.2927
		P(C)	0.2074	0.2074	0.2074	0.2074	0.2074	0.2073
		P(G)	0.2080	0.2080	0.2080	0.2081	0.2079	0.2080
12	133275309	P(A)	0.2957	0.2957	0.2958	0.2958	0.2956	0.2958
		P(T)	0.2966	0.2966	0.2966	0.2966	0.2967	0.2966
		P(C)	0.2035	0.2035	0.2035	0.2035	0.2035	0.2035
		P(G)	0.2042	0.2042	0.2041	0.2041	0.2042	0.2042
13	95789532	P(A)	0.3069	0.3068	0.3069	0.3069	0.3069	0.3068
		P(T)	0.3079	0.3078	0.3079	0.3079	0.3079	0.3078
		P(C)	0.1926	0.1927	0.1925	0.1926	0.1925	0.1927
		P(G)	0.1926	0.1927	0.1926	0.1926	0.1926	0.1927
14	87316725	P(A)	0.2945	0.2944	0.2946	0.2946	0.2946	0.2944
		P(T)	0.2970	0.2970	0.2970	0.2970	0.2970	0.2970
		P(C)	0.2040	0.2041	0.2039	0.2040	0.2039	0.2040
		P(G)	0.2045	0.2045	0.2045	0.2045	0.2045	0.2046
15	101991189	P(A)	0.2896	0.2896	0.2895	0.2897	0.2894	0.2896
		P(T)	0.2901	0.2901	0.2901	0.2900	0.2902	0.2900
		P(C)	0.2097	0.2097	0.2098	0.2098	0.2097	0.2097
		P(G)	0.2106	0.2106	0.2106	0.2105	0.2106	0.210
16	90338345	P(A)	0.2758	0.2757	0.2758	0.2758	0.2757	0.2758
		P(T)	0.2784	0.2785	0.2783	0.2784	0.2784	0.2784
		P(C)	0.2221	0.2222	0.2221	0.2221	0.2222	0.2222
		P(G)	0.2237	0.2236	0.2238	0.2237	0.2237	0.2237
17	83257441	P(A)	0.2730	0.2730	0.2730	0.2730	0.2730	0.2730
		P(T)	0.2738	0.2739	0.2737	0.2739	0.2738	0.2737
		P(C)	0.2258	0.2257	0.2259	0.2258	0.2257	0.2259
		P(G)	0.2273	0.2274	0.2273	0.2272	0.2274	0.2274
18	74792881	P(A)	0.3011	0.3012	0.3009	0.3011	0.3011	0.3010
		P(T)	0.3014	0.3013	0.3014	0.3014	0.3013	0.3014
		P(C)	0.1987	0.1985	0.1988	0.1987	0.1987	0.1986
		P(G)	0.1989	0.1989	0.1989	0.1988	0.1989	0.1990
19	58617616	P(A)	0.2591	0.2592	0.2590	0.2592	0.2591	0.2591
		P(T)	0.2615	0.2615	0.2615	0.2616	0.2616	0.2613
		P(C)	0.2388	0.2388	0.2388	0.2387	0.2388	0.2388
		P(G)	0.2406	0.2405	0.2407	0.2405	0.2405	0.2408
20	59605541	P(A)	0.2778	0.2779	0.2777	0.2778	0.2777	0.2779
		P(T)	0.2810	0.2808	0.2811	0.2810	0.2810	0.2809
		P(C)	0.2201	0.2202	0.2200	0.2201	0.2201	0.2201
		P(G)	0.2211	0.2212	0.2211	0.2211	0.2212	0.2211
21	33543332	P(A)	0.2964	0.2964	0.2965	0.2963	0.2964	0.2966
		P(T)	0.2946	0.2947	0.2945	0.2946	0.2946	0.2947
		P(C)	0.2047	0.2046	0.2047	0.2046	0.2049	0.2044
		P(G)	0.2043	0.2043	0.2043	0.2045	0.2041	0.2043
22	50818468	P(A)	0.2651	0.2651	0.2651	0.2651	0.2652	0.2651
		P(T)	0.2648	0.2648	0.2648	0.2648	0.2647	0.2650
		P(C)	0.2339	0.2339	0.2340	0.234	0.2341	0.2337
		P(G)	0.2361	0.2362	0.2360	0.2362	0.2360	0.2362
		P(A)	0.3019	0.3019	0.3018	0.3019	0.3018	0.3018

X	156040895	P(T)	0.3029	0.3029	0.3029	0.3029	0.3029	0.3029
		P(C)	0.1971	0.1970	0.1971	0.1971	0.1971	0.1971
		P(G)	0.1982	0.1982	0.1982	0.1982	0.1982	0.1982
Y	57227415	P(A)	0.2985	0.2986	0.2985	0.2985	0.2987	0.2984
		P(T)	0.3012	0.3013	0.3011	0.3013	0.3011	0.3012
		P(C)	0.2001	0.2001	0.2001	0.2001	0.2001	0.2001
		P(G)	0.2001	0.2000	0.2003	0.2001	0.2001	0.2003

It is surprising but for all 24 human chromosomes, each of rows of probabilities in Table 1 contains approximately the same value of probabilities of nucleotides in its 7 cells corresponding to the appropriate code-chain N_1 and its epi-chains $N_{2/1}$, $N_{2/2}$, $N_{3/1}$, $N_{3/2}$, $N_{3/3}$: small differences in the values of probabilities appear only in the fourth decimal place though each code-chain strongly differ from its epi-chains by its length and by its nucleotide sequence. It means that these epi-chains practically are copies (or duplicates) of their code-chains regarding probabilities of nucleotides A, T, C and G. In particular, the analogue of the second Chargaff's rule for DNA code-chains is executed in the epi-chains: $P(A) \approx P(T)$ and $P(C) \approx P(G)$.

But what one can say about probabilities of nucleotides in epi-chains of higher orders? Whether such equalities of probabilities also exist there? The author's study gives a positive answer on this question. Table 2 shows probabilities $P(A)$, $P(T)$, $P(C)$ and $P(G)$ in examples of epi-chains of the 10th order, the 50th order, the 100th order in the human chromosome № 1. Tabular data testify that in all shown epi-chains these probabilities have practically the same values like in their code-chain N_1 .

Table 2. Probabilities of nucleotides $P(A)$, $P(T)$, $P(C)$ and $P(G)$ in the human chromosome №1 for the following examples of randomly selected epi-chains: $N_{10/1}$ and $N_{10/9}$ – two epi-chains of the 10th order; $N_{50/1}$ and $N_{50/49}$ – two epi-chains of the 50th order; $N_{100/1}$ and $N_{100/99}$ – two epi-chains of the 100th order. N_1 refers to the code-chain of this chromosome (by analogy with Table 1).

	N_1	$N_{10/1}$	$N_{10/9}$	$N_{50/1}$	$N_{50/49}$	$N_{100/1}$	$N_{100/99}$
P(A)	0.2910	0.2910	0.2911	0.2910	0.2915	0.2910	0.2915
P(T)	0.2918	0.2918	0.2919	0.2917	0.2916	0.2917	0.2920
P(C)	0.2085	0.2084	0.2084	0.2086	0.2085	0.2086	0.2080
P(G)	0.2087	0.2088	0.2087	0.2088	0.2085	0.2087	0.2086

The author has received similar results of approximate equality of probabilities of nucleotides A, T, C and G in code-chains and their epi-chains of different orders in complete sets of chromosomes of a few model organisms, which are used long ago in the study of genetics, development and disease: a fruit fly *Drosophila melanogaster*, a nematode *Caenorhabditis elegans*, a house mouse *Mus musculus*, a plant *Arabidopsis thaliana*. Table 3 shows probabilities of nucleotides $P(A)$, $P(T)$, $P(C)$ and $P(G)$ in the complete set of nuclear chromosomes of *Arabidopsis thaliana*. These tabular data confirm that probabilities of nucleotides in the considered epi-chains practically repeat their probabilities in code-chains N_1 in all nuclear chromosomes.

Table 3. Probabilities $P(A)$, $P(T)$, $P(C)$ and $P(G)$ of nucleotides A, T, C and G in epi-chains of initial orders in all nuclear chromosomes of a plant *Arabidopsis thaliana*. Data relating epi-chains of the second order ($N_{2/1}$, $N_{2/2}$) and the third order ($N_{3/1}$, $N_{3/2}$, $N_{3/3}$) are presented jointly with data of code-chains N_1 . Two left columns show ordinary symbols of chromosomes and their length. Initial data relating to these chromosomes were accessed from the CenBank (<https://www.ncbi.nlm.nih.gov/genome/4>).

Chr	DNA length (bp)		N_1	$N_{2/1}$	$N_{2/2}$	$N_{3/1}$	$N_{3/2}$	$N_{3/3}$
1	30427671	P(A)	0.3208	0.3207	0.3209	0.3212	0.3206	0.3207
		P(T)	0.3204	0.3204	0.3204	0.3204	0.3207	0.3202
		P(C)	0.1796	0.1796	0.1796	0.1794	0.1798	0.1796

		P(G)	0.1791	0.1792	0.1791	0.1790	0.1789	0.1795
2	19698289	P(A)	0.3207	0.3205	0.3208	0.3205	0.3204	0.3210
		P(T)	0.3207	0.3208	0.3206	0.3210	0.3208	0.3203
		P(C)	0.1799	0.1800	0.1798	0.1797	0.1802	0.1797
		P(G)	0.1788	0.1787	0.1788	0.1788	0.1785	0.1790
3	23459830	P(A)	0.3191	0.3194	0.3189	0.3195	0.3192	0.3187
		P(T)	0.3176	0.3175	0.3177	0.3171	0.3179	0.3176
		P(C)	0.1816	0.1815	0.1816	0.1816	0.1813	0.1818
		P(G)	0.1817	0.1817	0.1818	0.1818	0.1816	0.1818
4	18585056	P(A)	0.3197	0.3197	0.3196	0.3198	0.3195	0.3197
		P(T)	0.3183	0.3182	0.3183	0.318	0.3183	0.3185
		P(C)	0.1814	0.1814	0.1815	0.1812	0.1817	0.1814
		P(G)	0.1806	0.1807	0.1805	0.1810	0.1805	0.1803
5	26975502	P(A)	0.3197	0.3197	0.3198	0.3196	0.3199	0.3198
		P(T)	0.3209	0.3208	0.3209	0.321	0.3206	0.3210
		P(C)	0.1792	0.1793	0.1791	0.1796	0.1790	0.1790
		P(G)	0.1802	0.1802	0.1802	0.1797	0.1805	0.1803

Till now only nuclear chromosomes were under consideration. But what about probabilities of nucleotides in epi-chains of mitochondrions, whose length much more shorter that length of nuclear chromosomes? Table 4 shows these probabilities for human mitochondrion whose length is only 16569 bp.

Table 4. Probabilities P(A), P(T), P(C) and P(G) of nucleotides A, T, C and G in epi-chains of initial orders in the human mitochondrion. Data relating epi-chains of the second order ($N_{2/1}$, $N_{2/2}$) and the third order ($N_{3/1}$, $N_{3/2}$, $N_{3/3}$) are presented jointly with data on the code-chain N_1 . Initial data relating to the human mitochondrion were accessed from https://www.ncbi.nlm.nih.gov/nuccore/NC_012920.1.

DNA length (bp)		N_1	$N_{2/1}$	$N_{2/2}$	$N_{3/1}$	$N_{3/2}$	$N_{3/3}$
16569	P(A)	0.3093	0.3095	0.3090	0.3147	0.3126	0.3006
	P(T)	0.2471	0.2496	0.2446	0.2433	0.2486	0.2493
	P(C)	0.3127	0.3087	0.3168	0.3111	0.3131	0.3140
	P(G)	0.1309	0.1322	0.1296	0.1309	0.1257	0.1362

These tabular data testify that in human mitochondrion its probabilities of nucleotides in the code-chain N_1 and in considered epi-chains are also approximately equal to each other but with a lower level of accuracy. The second Chargaff's rule is not executed at all in this short mitochondrion: P(A) is significantly different from P(T) and P(C) is significantly different from P(G).

Table 5 shows probabilities of nucleotides in a plant *Arabidopsis thailana* for its mitochondrion and its chloroplast. Tabular data testify that these probabilities in the considered epi-chains are approximately the same as in the code-chains N_1 . The lengths of the mitochondrion and the chloroplast significantly exceed 100,000 bp and are many times longer than the length of the human mitochondrion. The analogue of the second Chargaff's rule for nucleotide code-chains is executed in epi-chains of these DNA: $P(A) \approx P(T)$ and $P(C) \approx P(G)$.

Table 5. Probabilities P(A), P(T), P(C) and P(G) of nucleotides A, T, C and G in epi-chains of initial orders in the mitochondrion and the chloroplast of a plant *Arabidopsis thailana*. Data relating epi-chains of the second order ($N_{2/1}$, $N_{2/2}$) and the third order ($N_{3/1}$, $N_{3/2}$, $N_{3/3}$) are presented jointly with data of the code-chain N_1 . Initial data about the DNA were accessed from https://www.ncbi.nlm.nih.gov/nuccore/NC_037304.1 and https://www.ncbi.nlm.nih.gov/nuccore/NC_000932.1.

Chr	DNA length (bp)		N_1	$N_{2/1}$	$N_{2/2}$	$N_{3/1}$	$N_{3/2}$	$N_{3/3}$
MT	367808	P(A)	0.2793	0.2773	0.2812	0.2798	0.2766	0.2814
		P(T)	0.2731	0.2729	0.2732	0.2722	0.2758	0.2711
		P(C)	0.2253	0.2260	0.2246	0.2259	0.2245	0.2255
		P(G)	0.2224	0.2238	0.2211	0.2221	0.2231	0.2220
Chloro- plast	154478	P(A)	0.3143	0.3132	0.3153	0.3093	0.3201	0.3134
		P(T)	0.3228	0.3250	0.3207	0.3231	0.3242	0.3211
		P(C)	0.1845	0.1843	0.1846	0.1902	0.1777	0.1856
		P(G)	0.1785	0.1775	0.1794	0.1775	0.1781	0.1799

4. Nucleotide probabilities in epi-chains of prokaryotic genomes

What one can say about nucleotide probabilities in epi-chains of prokaryotic genomes? This Section represents author's results testified that in prokaryotic genomes the same situation relating these probabilities exists like in eukaryotic genomes (but with a slightly lower level of accuracy in equalities). Table 6 shows probabilities (or frequencies) of nucleotides A, T, C and G in all 19 genomes of bacteria and archaea from their full list in the article [Rapoport, Trifonov, 2012].

Table 6. Probabilities P(A), P(T), P(C) and P(G) of nucleotides A, T, C and G in epi-chains of initial orders in 19 prokaryotic genomes. Data relating epi-chains of the second order ($N_{2/1}$, $N_{2/2}$) and the third order ($N_{3/1}$, $N_{3/2}$, $N_{3/3}$) are presented jointly with data of code-chains N_1 .

№	Prokaryotes, DNA length		N_1	$N_{2/1}$	$N_{2/2}$	$N_{3/1}$	$N_{3/2}$	$N_{3/3}$
1	Aquifex aeolicus, 1551335 bp	P(A)	0.2841	0.2834	0.2849	0.2820	0.2859	0.2845
		P(T)	0.2811	0.2818	0.2805	0.2815	0.2827	0.2791
		P(C)	0.2168	0.2173	0.2163	0.2167	0.2156	0.2181
		P(G)	0.2179	0.2176	0.2183	0.2198	0.2157	0.2183
2	Acidobacteria bacterium, 4996384 bp	P(A)	0.2155	0.2156	0.2153	0.2165	0.2154	0.2146
		P(T)	0.2171	0.2169	0.2173	0.2167	0.2186	0.2161
		P(C)	0.2855	0.2855	0.2855	0.2836	0.2863	0.2868
		P(G)	0.2819	0.2819	0.2818	0.2833	0.2798	0.2825
3	Bradyrhizobium japonicum, 9224208 bp	P(A)	0.1819	0.1820	0.1818	0.1806	0.1836	0.1815
		P(T)	0.1815	0.1814	0.1816	0.1797	0.1825	0.1824
		P(C)	0.3184	0.3185	0.3183	0.3197	0.3166	0.3189
		P(G)	0.3182	0.3182	0.3182	0.3201	0.3173	0.3173
4	Bacillus subtilis, 4025326 bp	P(A)	0.2805	0.2801	0.2809	0.2813	0.2795	0.2807
		P(T)	0.2806	0.2806	0.2805	0.2809	0.2803	0.2805
		P(C)	0.2192	0.2196	0.2189	0.2181	0.2196	0.2200
		P(G)	0.2197	0.2197	0.2197	0.2196	0.2206	0.2189
5	Chlamydia trachomatis, 1025839 bp	P(A)	0.2942	0.2939	0.2945	0.2927	0.2963	0.2936
		P(T)	0.2930	0.2931	0.2930	0.2921	0.2935	0.2935
		P(C)	0.2067	0.2066	0.2067	0.2074	0.2049	0.2077
		P(G)	0.2061	0.2063	0.2058	0.2078	0.2053	0.2051
6	Chromobacterium violaceum, 127377 bp	P(A)	0.1670	0.1658	0.1682	0.1758	0.1651	0.1602
		P(T)	0.1819	0.1812	0.1827	0.1883	0.1808	0.1766
		P(C)	0.3049	0.3044	0.3055	0.3009	0.3112	0.3027
		P(G)	0.3461	0.3486	0.3437	0.3350	0.3429	0.3606
7	Dehalococcoides mccartyi, 1521287 bp	P(A)	0.2664	0.2665	0.2663	0.2706	0.2617	0.2668
		P(T)	0.2645	0.2647	0.2643	0.2653	0.2686	0.2596
		P(C)	0.2338	0.2337	0.2339	0.2295	0.2376	0.2343
		P(G)	0.2353	0.2351	0.2355	0.2346	0.2321	0.2394

8	Escherichia coli, 5231428 bp	P(A)	0.2480	0.2481	0.248	0.2479	0.2468	0.2494
		P(T)	0.2472	0.2470	0.2474	0.2489	0.2461	0.2466
		P(C)	0.2526	0.2523	0.2528	0.2519	0.2544	0.2515
		P(G)	0.2522	0.2526	0.2518	0.2514	0.2527	0.2525
9	Flavobacterium psychrophilum, 2860382 bp	P(A)	0.3326	0.3326	0.3326	0.3335	0.3306	0.3338
		P(T)	0.342	0.3416	0.3425	0.3411	0.3423	0.3427
		P(C)	0.1640	0.1643	0.1637	0.1616	0.1665	0.1638
		P(G)	0.1614	0.1615	0.1612	0.1638	0.1606	0.1597
10	Gloeobacter violaceus, 4659019 bp	P(A)	0.1906	0.1905	0.1907	0.1892	0.1925	0.1901
		P(T)	0.1894	0.1899	0.1890	0.1881	0.1899	0.1903
		P(C)	0.3101	0.3098	0.3104	0.312	0.3074	0.3108
		P(G)	0.3099	0.3099	0.3099	0.3107	0.3103	0.3088
11	Helicobacter pylori, 1643831 bp	P(A)	0.3033	0.3034	0.3032	0.3033	0.3042	0.3022
		P(T)	0.3048	0.3049	0.3048	0.3037	0.3028	0.3080
		P(C)	0.1970	0.1969	0.1970	0.1989	0.1958	0.1962
		P(G)	0.1949	0.1949	0.1950	0.1941	0.1972	0.1936
12	Methanosarcina acetivorans, 5751492 bp	P(A)	0.2848	0.2848	0.2848	0.2830	0.2851	0.2863
		P(T)	0.2884	0.2881	0.2887	0.2890	0.2875	0.2887
		P(C)	0.2136	0.2139	0.2132	0.2148	0.2133	0.2126
		P(G)	0.2132	0.2132	0.2133	0.2132	0.2141	0.2124
13	Nanoarchaeum equitans, 490885 bp	P(A)	0.3422	0.3424	0.3420	0.3413	0.3423	0.3431
		P(T)	0.3422	0.3419	0.3425	0.342	0.3413	0.3433
		P(C)	0.1576	0.1581	0.1571	0.1593	0.1571	0.1563
		P(G)	0.1580	0.1577	0.1583	0.1573	0.1594	0.1573
14	Syntrophus aciditrophicus, 3179300 bp	P(A)	0.2431	0.2431	0.2431	0.2441	0.2418	0.2433
		P(T)	0.2423	0.2423	0.2424	0.2446	0.2413	0.2412
		P(C)	0.2556	0.2556	0.2557	0.2546	0.2573	0.2551
		P(G)	0.2590	0.2590	0.2589	0.2567	0.2597	0.2604
15	Streptomyces coelicolor, 8667507 bp	P(A)	0.1389	0.1386	0.1391	0.1366	0.1408	0.1392
		P(T)	0.1400	0.1400	0.1399	0.1381	0.1414	0.1403
		P(C)	0.3601	0.3600	0.3602	0.3614	0.3580	0.3609
		P(G)	0.3611	0.3613	0.3608	0.3638	0.3598	0.3596
16	Sulfolobus solfataricus, 2727337 bp	P(A)	0.3181	0.3179	0.3184	0.3182	0.3185	0.3177
		P(T)	0.3233	0.3231	0.3234	0.3225	0.3230	0.3243
		P(C)	0.1798	0.1797	0.1799	0.1798	0.1782	0.1814
		P(G)	0.1788	0.1793	0.1783	0.1795	0.1803	0.1765
17	Treponema denticola, 1850823 bp	P(A)	0.3095	0.3095	0.3096	0.3079	0.3098	0.3109
		P(T)	0.3106	0.3104	0.3107	0.3120	0.3089	0.3108
		P(C)	0.188	0.1877	0.1882	0.1882	0.1879	0.1877
		P(G)	0.1919	0.1923	0.1915	0.1918	0.1934	0.1906
18	Thermotoga maritima, 1859582 bp	P(A)	0.2695	0.2700	0.2689	0.2703	0.2697	0.2684
		P(T)	0.2678	0.2681	0.2675	0.2689	0.2694	0.2651
		P(C)	0.2281	0.2277	0.2284	0.2271	0.2289	0.2282
		P(G)	0.2347	0.2341	0.2352	0.2337	0.2320	0.2383
19	Thermus thermophilus, 2121526 bp	P(A)	0.1543	0.1543	0.1542	0.1515	0.1542	0.1571
		P(T)	0.1557	0.1556	0.1558	0.1556	0.1529	0.1586
		P(C)	0.3461	0.3458	0.3464	0.3490	0.3465	0.3428
		P(G)	0.3439	0.3443	0.3435	0.3439	0.3464	0.3414

The following genomes are shown in Table 2 under numbers 1-19:

- 1) *Aquifex aeolicus* VF5, complete genome, 1551335 bp, accession AE000657, version AE000657.1, <https://www.ncbi.nlm.nih.gov/nucleotide/AE000657.1?report=genbank> ;
- 2) *Acidobacteria bacterium* KBS 146 M015DRAFT_scf7180000000004_quiver.1_C, whole genome shotgun sequence, 4996384 bp, accession JHVA01000001, <https://www.ncbi.nlm.nih.gov/nucleotide/JHVA01000001.1?report=genbank>;
- 3) *Bradyrhizobium japonicum* strain E109, complete genome, 9224208 bp, accession CP010313, <https://www.ncbi.nlm.nih.gov/nucleotide/CP010313.1?report=genbank> ;
- 4) *Bacillus subtilis* strain UD1022, complete genome, 4025326 bp, accession CP011534, <https://www.ncbi.nlm.nih.gov/nucleotide/CP011534.1?report=genbank>;
- 5) *Chlamydia trachomatis* strain QH111L, complete genome, 1025839 bp, accession CP018052, <https://www.ncbi.nlm.nih.gov/nucleotide/CP018052.1?report=genbank>;
- 6) *Chromobacterium violaceum* strain LK30 1, whole genome shotgun sequence, 127377 bp, accession LDUX01000001 version LDUX01000001.1, <https://www.ncbi.nlm.nih.gov/nucleotide/LDUX01000001.1?report=genbank>;
- 7) *Dehalococcoides mccartyi* strain CG3, complete genome, NCBI Reference Sequence: NZ_CP013074.1, 1521287 bp, https://www.ncbi.nlm.nih.gov/nucleotide/NZ_CP013074.1?report=genbank;
- 8) *Escherichia coli* CFT073, complete genome, GenBank: AE014075.1, 5231428 bp, <https://www.ncbi.nlm.nih.gov/nucleotide/AE014075.1?report=genbank>;
- 9) *Flavobacterium psychrophilum* JIP02/86, complete genome, 2860382 bp, accession NC_009613, https://www.ncbi.nlm.nih.gov/nucleotide/NC_009613.3;
- 10) *Gloeobacter violaceus* PCC 7421 DNA, complete genome, GenBank: BA000045.2, 4659019 bp, accession BA000045 AP006568-AP006583 version BA000045.2, <https://www.ncbi.nlm.nih.gov/nucleotide/BA000045.2?report=genbank>;
- 11) *Helicobacter pilory*, NCBI Reference Sequence: NC_000921.1, complete genome, 1643831 bp, accession NC_000921 NZ_AE001440-NZ_AE001571 version NC_000921.1, https://www.ncbi.nlm.nih.gov/nucleotide/NC_000921.1 ;
- 12) *Methanosarcina acetivorans* str. C2A, complete genome, 5751492 bp, accession AE010299 AE010656-AE011189 version AE010299.1, <https://www.ncbi.nlm.nih.gov/nucleotide/AE010299>;
- 13) *Nanoarchaeum equitans* Kin4-M, complete genome, 490885 bp, accession AE017199 AACL01000000 AACL01000001 version AE017199.1, <https://www.ncbi.nlm.nih.gov/nucleotide/AE017199.1?report=genbank>;
- 14) *Syntrophus aciditrophicus* SB, complete genome, 3179300 bp, accession CP000252, <https://www.ncbi.nlm.nih.gov/nucleotide/CP000252.1?report=genbank>;
- 15) *Streptomyces coelicolor* A3(2) complete genome, 8667507 bp, accession AL645882, <https://www.ncbi.nlm.nih.gov/nucleotide/AL645882.2?report=genbank>;
- 16) *Sulfolobus solfataricus* strain SULA, complete genome, 2727337 bp, accession CP011057, <https://www.ncbi.nlm.nih.gov/nucleotide/CP011057.1?report=genbank>;
- 17) *Treponema denticola* SP33 supercont1.1, whole genome shotgun sequence, NCBI Reference Sequence: NZ_KB442453.1, 1850823 bp, accession NZ_KB442453 NZ_AGZ01000000 version NZ_KB442453.1, https://www.ncbi.nlm.nih.gov/nucleotide/NZ_KB442453.1?report=genbank
- 18) *Thermotoga maritima* strain Tma200, complete genome, 1859582 bp, accession CP010967, <https://www.ncbi.nlm.nih.gov/nucleotide/CP010967.1?report=genbank>;
- 19) *Thermus thermophilus* DNA, complete genome, strain: TMY, 2121526 bp, accession AP017920, <https://www.ncbi.nlm.nih.gov/nucleotide/AP017920.1?report=genbank>

Data in Table 6 testify that in prokaryotic genomes probabilities of nucleotides A, T, C and G in considered DNA epi-chains are approximately repeated their probabilities in code-chains like the case of eukaryotic genomes considered above.

5. Regarding probabilities of triplets in DNA epi-chains of different orders

This Section presents initial author's results on studying probabilities of each of 64 triplets in long DNA epi-chains of various orders. These results testify in favor that in relation to triplets all epi-chains of each concrete order have practically identical probabilities, which can be used as an individual characteristic of the order of epi-chains (in contrast to the above described situation with nucleotide probabilities, which practically identical in epi-chains of different orders having no individual characteristics from this standpoint). Such differences in individual probability characteristics of various orders of epi-chains allow speculating on using these peculiarities of various epi-chains sets in genetic processes.

Table 7 shows probabilities (or frequencies) of each of 64 triplets in epi-chains of various orders in the human chromosome № 1 having 248956422 bp.

Table 7. Probabilities of 64 triplets in epi-chains of human chromosome № 1, which contains 248956422 bp. The cases of epi-chains of the second order ($N_{2/1}$, $N_{2/2}$), the third order ($N_{3/1}$, $N_{3/2}$, $N_{3/3}$) and the fourth order ($N_{4/1}$, $N_{4/2}$, $N_{4/3}$, $N_{4/4}$) are shown. All values are rounded to four decimal places. Epi-chains of each of orders are marked by an individual color. N_1 refers to the code-chain of this chromosome. Initial data relating to the chromosome were accessed from https://www.ncbi.nlm.nih.gov/nucore/NC_000001.11.

	N_1	$N_{2/1}$	$N_{2/2}$	$N_{3/1}$	$N_{3/2}$	$N_{3/3}$	$N_{4/1}$	$N_{4/2}$	$N_{4/3}$	$N_{4/4}$
CCC	0.0138	0.0122	0.0122	0.0138	0.0138	0.0138	0.0124	0.0124	0.0124	0.0124
CCG	0.0029	0.0123	0.0123	0.0103	0.0103	0.0103	0.0109	0.0110	0.0110	0.0110
CCT	0.0185	0.0131	0.0131	0.0126	0.0127	0.0127	0.0135	0.0134	0.0134	0.0134
CCA	0.0188	0.0116	0.0116	0.0121	0.0121	0.0121	0.0119	0.0119	0.0119	0.0119
CGC	0.0025	0.0113	0.0113	0.0107	0.0107	0.0107	0.0117	0.0117	0.0117	0.0117
CGG	0.0029	0.0122	0.0122	0.0102	0.0103	0.0103	0.0109	0.0109	0.0109	0.0109
CGT	0.0026	0.0111	0.0111	0.0133	0.0133	0.0133	0.0123	0.0124	0.0123	0.0124
CGA	0.0023	0.0132	0.0132	0.0122	0.0121	0.0122	0.0125	0.0125	0.0125	0.0125
CTC	0.0176	0.0139	0.0140	0.0137	0.0137	0.0137	0.0130	0.0130	0.0130	0.0129
CTG	0.0209	0.0116	0.0116	0.0110	0.0110	0.0110	0.0103	0.0104	0.0103	0.0103
CTT	0.0201	0.0186	0.0186	0.0175	0.0175	0.0175	0.0180	0.0180	0.0180	0.0180
CTA	0.0127	0.0149	0.0150	0.0149	0.0149	0.0149	0.0153	0.0152	0.0152	0.0152
CAC	0.0152	0.0112	0.0113	0.0127	0.0127	0.0127	0.0130	0.0130	0.0130	0.0130
CAG	0.0210	0.0114	0.0114	0.0108	0.0109	0.0109	0.0102	0.0102	0.0103	0.0102
CAT	0.0179	0.0137	0.0137	0.0164	0.0164	0.0164	0.0161	0.0162	0.0161	0.0162
CAA	0.0186	0.0161	0.0161	0.0161	0.0162	0.0162	0.0165	0.0165	0.0164	0.0164
GCC	0.0125	0.0109	0.0109	0.0094	0.0094	0.0094	0.0104	0.0104	0.0104	0.0104
GCG	0.0025	0.0114	0.0114	0.0107	0.0107	0.0107	0.0116	0.0116	0.0116	0.0116
GCT	0.0144	0.0116	0.0116	0.0106	0.0105	0.0106	0.0108	0.0108	0.0108	0.0108
GCA	0.0146	0.0104	0.0104	0.0105	0.0105	0.0105	0.0116	0.0116	0.0117	0.0117
GGC	0.0126	0.0109	0.011	0.0093	0.0094	0.0093	0.0104	0.0104	0.0103	0.0103
GGG	0.0138	0.0121	0.0121	0.0137	0.0137	0.0137	0.0124	0.0123	0.0124	0.0124
GGT	0.0119	0.0129	0.0129	0.012	0.0119	0.0119	0.0129	0.0129	0.0129	0.0129
GGA	0.016	0.0133	0.0133	0.0139	0.0138	0.0138	0.0131	0.0131	0.0131	0.0131
GTC	0.0096	0.0112	0.0112	0.0122	0.0122	0.0122	0.0116	0.0116	0.0117	0.0116
GTG	0.0153	0.0114	0.0114	0.0128	0.0128	0.0127	0.0132	0.0132	0.0131	0.0132
GTT	0.0145	0.0165	0.0165	0.0162	0.0162	0.0162	0.0165	0.0165	0.0165	0.0165
GTA	0.0112	0.0149	0.0148	0.0167	0.0166	0.0167	0.0165	0.0166	0.0165	0.0165
GAC	0.0096	0.0114	0.0114	0.0123	0.0123	0.0123	0.0116	0.0115	0.0115	0.0116
GAG	0.0176	0.0138	0.0138	0.0139	0.0138	0.0138	0.013	0.0131	0.0131	0.0130
GAT	0.0133	0.0168	0.0168	0.0162	0.0162	0.0162	0.0160	0.0160	0.0160	0.0160
GAA	0.0196	0.0193	0.0193	0.0186	0.0186	0.0186	0.0174	0.0173	0.0173	0.0173
TCC	0.0159	0.0133	0.0132	0.0138	0.0138	0.0138	0.0131	0.0131	0.0131	0.0131
TCG	0.0023	0.0130	0.0130	0.0122	0.0122	0.0122	0.0124	0.0124	0.0125	0.0124

TCT	0.0223	0.0190	0.0191	0.0183	0.0183	0.0183	0.0170	0.0170	0.0170	0.0171
TCA	0.0196	0.0158	0.0159	0.0164	0.0164	0.0164	0.0154	0.0154	0.0154	0.0154
TGC	0.0146	0.0104	0.0104	0.0107	0.0107	0.0107	0.0117	0.0117	0.0117	0.0116
TGG	0.019	0.0118	0.0118	0.0122	0.0122	0.0122	0.0120	0.0120	0.0120	0.0120
TGT	0.0199	0.0147	0.0147	0.0171	0.0171	0.0171	0.0171	0.0171	0.0170	0.0171
TGA	0.0195	0.0160	0.0160	0.0166	0.0166	0.0166	0.0154	0.0154	0.0154	0.0153
TTC	0.0197	0.0193	0.0193	0.0185	0.0184	0.0185	0.0174	0.0174	0.0174	0.0174
TTG	0.0188	0.0162	0.0162	0.0163	0.0163	0.0163	0.0166	0.0166	0.0166	0.0166
TTT	0.0372	0.0374	0.0374	0.0321	0.0321	0.0322	0.0340	0.0341	0.0340	0.0341
TTA	0.0198	0.0258	0.0258	0.0249	0.0249	0.025	0.0263	0.0263	0.0263	0.0263
TAC	0.011	0.0149	0.0149	0.0167	0.0167	0.0167	0.0166	0.0166	0.0166	0.0166
TAG	0.0128	0.0150	0.0150	0.0149	0.0149	0.0149	0.0152	0.0152	0.0152	0.0152
TAT	0.0194	0.0231	0.0231	0.0262	0.0263	0.0263	0.0253	0.0253	0.0254	0.0254
TAA	0.0199	0.0257	0.0257	0.0248	0.0249	0.0248	0.0262	0.0263	0.0262	0.0262
ACC	0.0118	0.0128	0.0128	0.0118	0.0118	0.0118	0.0129	0.0129	0.0128	0.0129
ACG	0.0025	0.0111	0.0111	0.0133	0.0133	0.0133	0.0125	0.0125	0.0125	0.0124
ACT	0.0162	0.0153	0.0153	0.0155	0.0156	0.0156	0.0153	0.0153	0.0153	0.0152
ACA	0.0198	0.0146	0.0146	0.0172	0.0172	0.0172	0.0168	0.0169	0.0169	0.0168
AGC	0.0144	0.0116	0.0115	0.0104	0.0104	0.0104	0.0108	0.0108	0.0107	0.0108
AGG	0.0185	0.0131	0.0131	0.0127	0.0127	0.0126	0.0134	0.0133	0.0134	0.0133
AGT	0.0161	0.0152	0.0152	0.0155	0.0156	0.0155	0.0154	0.0154	0.0154	0.0155
AGA	0.0224	0.0188	0.0188	0.0183	0.0183	0.0183	0.0169	0.0170	0.0169	0.0169
ATC	0.0132	0.0168	0.0167	0.0163	0.0163	0.0163	0.0159	0.0159	0.0159	0.0159
ATG	0.0178	0.0138	0.0137	0.0165	0.0165	0.0165	0.0160	0.0160	0.0161	0.0160
ATT	0.0239	0.0263	0.0262	0.0259	0.0260	0.0261	0.0258	0.0258	0.0259	0.0259
ATA	0.0194	0.0232	0.0231	0.0262	0.0262	0.0262	0.0254	0.0253	0.0254	0.0254
AAC	0.0145	0.0163	0.0164	0.0162	0.0162	0.0162	0.0163	0.0163	0.0163	0.0163
AAG	0.0199	0.0185	0.0185	0.0173	0.0173	0.0172	0.0180	0.0180	0.0180	0.0180
AAT	0.0238	0.0264	0.0263	0.0260	0.0260	0.0260	0.0257	0.0257	0.0257	0.0257
AAA	0.0369	0.0373	0.0373	0.0318	0.0317	0.0318	0.0339	0.0339	0.0339	0.0339

These tabular data show, for example, that probabilities $P(\text{CCC})$ of the triplet CCC are different in the code-chain N_1 and in epi-chains of various order but practically identical in all epi-chains of the same order:

- In the code-chain N_1 , $P(\text{CCC}) = 0,0138$;
- In all epi-chains of the second order $N_{2/1}$ and $N_{2/2}$, $P(\text{CCC}) = 0,0122$;
- In all epi-chains of the third order $N_{3/1}, N_{3/2}$ and $N_{3/3}$, $P(\text{CCC}) = 0,0138$;
- In all epi-chains of the fourth order $N_{4/1}, N_{4/2}, N_{4/3}$ and $N_{4/4}$, $P(\text{CCC}) = 0,0124$.

Probabilities of all other triplets in Table 7 show a similar strong connection with orders of epi-chains; inside all epi-chains of the same order, these probabilities have only small fluctuations in the fourth decimal place.

6. Regarding a supposed genetic role of epi-chains

This Section briefly describes author's thoughts about described numerical phenomena and about a possible genetic role of DNA epi-chains.

The above presented results testify in favor of existence of the following **two universal genetic rules relating to epi-chains in eukariotic and prokariotic genomes**:

- in long single-stranded DNA sequences of eucariotic and prokariotic genomes, probabilities of nucleotides A, T, C and G in their code-chains are practically repeated in their epi-chains of n -th orders ($n = 2, 3, 4, 5, \dots$ but not extremely large);
- in long DNA epi-chains, probabilities of nucleotides A and T are approximately equal ($P(A) \approx P(T)$) and probabilities of nucleotides C and G are also approximately equal ($P(C) \approx P(G)$) (this rule for epi-chains is analogous to the second Chargaff's rule for DNA code-chains).

Of course, further researches are needed to define a degree of unicursality of these epi-chains rules.

These rules belong to the field of quantum biology and support the Jordan's idea that *«life's missing laws were the rules of chance and probability (the indeterminism) of the quantum world that were somehow scaled up inside living organisms»* [McFadden, Al-Khalili, 2018]. These rules tell about probabilities, which is the main term in the language of quantum mechanics and quantum informatics.

The presented results demonstrate that long DNA sequences are not random at all but they are regular constructions with their rich nets (or lattices) of epi-chains interrelated with each other and with code-chains. Integer numbers $n = 1, 2, 3, 4, \dots$ play essential role in these regular constructions, which can be termed as harmonical in this relation. The arrangement of nucleotides in long DNA is not at all arbitrary, but organized in such a way that the rules for nucleotide probabilities are implemented on the DNA code-chain and the wide set of DNA epi-chains. Long DNA sequences demonstrate themselves as quantum-mechanical entities, in which there is a long-range interaction between the elements. Quantum entanglement in this quantum genetic system can be essential for its surprisingly organized construction. The thought about an important role of quantum entanglement in organization of DNA is not new: see, for example, the article *«Quantum entanglement between the electron clouds of nucleic acids in DNA»* about the entanglement as a glue for DNA constructions where lattice vibrations or phonons are significant [Rieper, Anders, Vedral, 2011].

Each DNA epi-chain is a complex vibration system with many degrees of freedom and correspondingly with many resonance frequencies. Energetic states and sets of resonance frequencies of DNA epi-chains can be changed under influence of external physical and chemical factors, for example, under influence of joining new chemical elements or bonds to certain sites of DNA or under DNA supercoiling. Since nucleotide code-chains on DNA strands carry genetic information, the phenomenon of numerical equalities in the probabilities of nucleotides between code-chains and epi-chains provokes the following important thought: **not only DNA code-chains but also epi-chains and their nucleotide sequences are participants of genetic and epigenetic processes and play an informational genetic and epigenetic role.**

The author thinks that such role of epi-chains can be connected with their physical essence as vibration systems with many resonance frequencies. Concerning importance of resonances in quantum mechanical systems, E. Schrodinger wrote: *“The one thing which one has to accept and which is the inalienable consequence of the wave-equation as it is used in every problem, under the most various forms, is this: that the interaction between two microscopic physical systems is controlled by a peculiar law of resonance»* [Schrodinger, 1952, p.115]. Structures of molecular genetics are deeply connected with mathematical formalisms of the theory of resonances of vibration systems with many degrees of freedom [Petoukhov, 2015a,b,c, 2016a; Petoukhov, Petukhova, 2017]. From the point of view of quantum mechanics, the interaction of molecules is based on the emission and absorption of photons with the participation of resonance correspondences. The energy state of parts of molecular-genetic systems depends on the emission and absorption of photons, which are the force carriers for electromagnetic field (see some details in [Hu, Petoukhov, Petukhova, 2018]). The creator of the theory of resonances in structural chemistry L. Pauling stated about an important meaning of resonances in organization of living matter [Pauling, 1940]. The book [Ji, 2017] contains new data about a role of resonances in cells.

In various cells, energetic states and sets of resonance frequencies of their DNA epi-chains can be mechanically depended on a location of separate cells inside a holistic ensemble of cells. Such differences in energetic states and resonance frequencies of DNA epi-chains in various cells can explain the phenomenon of different fate of various cells in ontogenesis, although all cells contain DNA of the same nucleotide composition. This also relates the well-known embryological rule: *“The fate of an embryo part is a function of its position within a whole”* [Belousov, 2015, p.4; Driesch, 1921].

The additional argument in favor of a participation of DNA epi-chains in epigenetic phenomena is given by the results described in the previous Section (Table 7) about regular probabilities of triplets in long DNA epi-chains of various orders. The author believes that knowledge on DNA epi-chains and on DNA telomeres allows developing new model versions of senescence. For example, one of possible model approaches considers epi-chains starting at the most distal nucleotides of telomeres. The model assumes that at this stage of cell life, the epi-chain of DNA has the predominant epigenetic activity, which contains this most distal nucleotide №1 and has the smallest order among all epi-chains containing this telomeric nucleotide №1. Under cell division, the telomeric ends become shorter and the nucleotide №1 is removed, giving the most distal place in DNA to another nucleotide

№2. At this new stage of cell life, a new epi-chain gets the predominant epigenetic activity, which contains this telomeric nucleotide №2 and has the smallest order among all epi-chains containing this telomeric nucleotide. This relay process of transferring activity to new epi-chains, which have specific energetic states and resonance frequencies, repeats with each cell division. More extended explanation of this senescence model, which combines knowledge about DNA telomeres and epi-chains, will be published some later.

In addition, DNA epi-chains with their helix configurations and changeable energetic states seem to participate in DNA supercoiling. Supercoiling, which needs an additional energy, is important in a number of biological processes, such as compacting DNA, DNA metabolism and possibly gene expression (https://en.wikipedia.org/wiki/DNA_supercoil). Changing energetic states of epi-chains, which is connected with changing their sets of resonance frequencies, can serve as a manage factor for DNA supercoiling.

The author's study of DNA epi-chains provokes many new questions for further investigations, for example, the following: do epi-chains with their own nucleotide sequences have their own epi-codes or not? Are there nucleotide sequences on DNA epi-chains that correspond to genes on the DNA code-chain? What will the comparative analysis of the characteristics of epi-chains in genomes of different organisms for the tasks of evolutionary biology?

This article presents the author's results about nucleotide probabilities relating to epi-chains of single-stranded DNA. But nucleotide probabilities on epi-chains of both strands of DNA double helix should be also studied. Such epi-chains should consist of a sequence of numbered nucleotides taken alternately from each of the two code-chains of DNA. For a preliminary description of such epi-chains of DNA double helix, one can denote by the symbol N_{1a} the code-chain of the first DNA strand and by the symbol N_{1b} the code-chain of the second DNA strand. Fig. 3 shows fragments of two code-chains N_{1a} and N_{1b} of a DNA double helix with an indication of epi-chains of the second and the third orders. For example, the epi-chain of the second order, which begins from the nucleotide 1a, contains the nucleotide sequence 1a-3b-5a-7b-9a-11b-... . The epi-chain of the third order, which begins from the nucleotide 1b, contains the nucleotide sequence 1b-4a-7b-10a-13b-... .

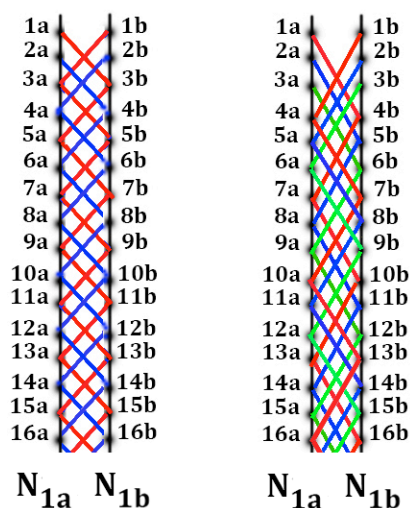


Fig. 3. The conditional presentation of two code-chains N_{1a} and N_{1b} of a DNA double helix in a form of straight lines carrying numerated nucleotides. Left: the system of epi-chains of the second order is shown, which begin from nucleotides 1a and 1b (red lines) and from nucleotides 2a and 2b (blue lines). Right: the system of epi-chains of the third order is shown, which begin from nucleotides 1a and 1b (red lines), from nucleotides 2a and 2b (blue lines) and from 3a and 3b (green lines).

The study of the nucleotide probabilities on epi-chains of DNA double helices is being conducted by the author at the present time, and their results is planed to publish later. The described studies conducted by the author use computer programs developed by his graduate student V.I. Svirin according to the author's technical task.

7. 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]. Many authors have supposed that living organisms use principles of quantum informatics [Abbott, Davies, Pati, 2008; Altaisky, Filatov, 2001; Fimmel, Petoukhov, 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].

The author's results presented in this article support Jordan's ideas of quantum biology by means of study of probabilities of nucleotides located on special helix trajectories in DNA strands.

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 author is grateful to G. Darvas, E. Fimmel, M. He, Z.B. Hu, Yu.I. Manin and I.V. Stepanyan for their collaboration. Special thanks to V.I. Svirin who created computer programs for described researches on the basis of author's technical tasks. Special thanks also to the German Academic Exchange Service (DAAD) for providing the very useful internship for the author in autumn 2017 at the Institute of Mathematical Biology of the Mannheim University of Applied Sciences (Germany) where the host was Prof. E. Fimmel.

References

- Abbott D., Davies P.C.W., Pati A.K.** (Eds.) (2008). *Quantum Aspects of Life*, foreword by Sir Roger Penrose, 2008 / ISBN-13: 978-1-84816-253-2.
- Albrecht-Buehler G.** (2006). Asymptotically increasing compliance of genomes with Chargaff's second parity rules through inversions and inverted transpositions. - *Proc Natl Acad Sci USA*. 103 (47): p. 17828–17833, doi:10.1073/pnas.0605553103.
- Altaisky M.V., Filatov F.P.**, 2001. Genetic information and quantum gas. - arXiv:quant-ph/0106123v1, submitted on 22.06.2001.
- Belousov L.V.** (2015). *Morphomechanics of development*. Springer International Publishing Switzerland, ISBN 978-3-319-13989-0, DOI 10.1007/978-3-319-13990-6.
- Chargaff E., Lipshitz R., Green C.** (1952). Composition of the deoxyribose nucleic acids of four genera of searichin. - *J Biol Chem*. 195 (1), p. 155–160.
- Chargaff E.** (1971). Preface to a Grammar of Biology: A hundred years of nucleic acid research, *Science*, 172, p. 637-642.
- Darvas G.** (2018). Petoukhov's rules on symmetries in long DNA-texts. *Symmetry: Culture and Science*, v. 29, # 2, pages 318-320, https://doi.org/10.26830/symmetry_2018_2_318, <http://journal-scs.symmetry.hu/abstract/?pid=673>
- Driesch H.** (1921). *Philosophie des Organischen*. Engelmann, Leipzig.

- 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.
- 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. doi:10.1007/s11224-016-0880-0, <http://link.springer.com/journal/11224/28/1/page/2>
- Hu Z.B., Petoukhov S.V., Petukhova E.S.** (2017b). I-Ching, dyadic groups of binary numbers and the geno-logic coding in living bodies. - *Progress in Biophysics and Molecular Biology*, v. 131, December 2017, pp. 354-368. (<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.U.** (1993). Quantum mechanical properties of biosystems: a framework for complexity, structural stability, and transformations.- *Biosystems*, 31 (1), pp. 65–73.
- Igamberdiev A.I.**, 2004. Quantum computation, non-demolition measurements, and reflective control in living systems. – *BioSystems*, v. 77, p. 47–56.
- Igamberdiev A.I.**, 2007. Physical limits of computation and emergence of life. – *BioSystems*, v. 90, p. 340–349.
- Igamberdiev A.I.**, 2008. Objective patterns in the evolving network of non-equivalent observers. – *BioSystems*, v. 92, p. 122–131.
- Igamberdiev A.I., Shklovskiy-Kordi N.E.**, 2016. Computational power and generative capacity of genetic systems. – *BioSystems*, v. 142–143, p. 1–8.
- Igamberdiev A.I., Shklovskiy-Kordi N.E.**, 2017. The quantum basis of spatiotemporality in perception and consciousness. - *Progress in Biophysics and Molecular Biology*, v. 130, p. 15-25.
- Ji S.** (2017). *The Cell Language Theory: Connecting Mind and Matter*. - World Scientific Publishing Europe Ltd. , 450 p. ISBN-13: 978-1848166608.
- Jordan P.** (1932). Die Quantenmechanik und die Grundprobleme der Biologie und Psychologie. *Naturwissenschaften* 20, 815–821. (doi:10.1007/BF01494844).
- Josephson B.D.** (2018). The Physics of Mind and Thought. Preprint, DOI: 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, pp. 15–19.
- Matsuno K.** (2003). Quantum mechanics in first, second and third person descriptions. – *BioSystems*, 68, pp. 107-118.
- Matsuno K., Paton R.C.** (2000). Is there a biology of quantum information? – *BioSystems*, 55, p.39–46.
- McFadden J., Al-Khalili J.** (2018). The origins of quantum biology. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*. 12 December 2018. <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.
- Patel A.** (2001a). Quantum algorithms and the genetic code. *Pramana – Journal of Physics*, v. 56, 2-3, pp. 367-381, arXiv:quant-ph/0002037.
- Patel A.** (2001b). Testing quantum dynamics in genetic information processing. – *Journal of Genetics*, v. 80, 1, pp. 39-43.
- Patel A.** (2001c). Why genetic information processing could have a quantum basis. – *Journal of Biosciences*, v. 26, 2, pp. 145-151.
- Pauling L.** (1940). *The Nature of the Chemical Bond and the Structure of Molecules and Crystals: An Introduction to Modern Structural Chemistry*, 2nd ed. Oxford University Press, London, 664 p.
- Penrose R.** (1996). *Shadows of the Mind: A Search for the Missing Science of Consciousness*. - Oxford University Press, USA, 480 p.
- Penrose R.** (2019). Your eyes are not meant for seeing. – *Community*, February 21, 2019, <https://thriveglobal.com/stories/your-eyes-are-not-meant-for-seeing/> .
- Petoukhov S.V.** (2015a). Resonances and genetic biomechanics. - *Symmetry: Culture and Science*, т. 26, №3, p. 379-397, http://petoukhov.com/PETOUKHOV_IN_SCS_2015.pdf
- Petoukhov S.V.** (2015b). The concept of resonances in genetics ("resonant bioinformatics"). - Selected Works of the VII International Congress "Weak and Superweak Fields and Radiation in Biology and Medicine", September 07-11, 2015, St. Petersburg, pp. 1-41,

<http://www.biophys.ru/lib/sci/resonance/460-resonance-00001> (in Russian).

- Petoukhov S. V.** (2015c). Music and the Modeling Approach to Genetic Systems of Biological Resonances. *Presentation at the conference «IS4IS Summit Vienna 2015. The information society at the crossroads»*, 3-7 June 2015, Vienna, Austria, <http://sciforum.net/conference/70/paper/2812>
- Petoukhov S.V.** (2016a). The system-resonance approach in modeling genetic structures. – *Biosystems*, January 2016, v. 139, p. 1-11, http://petoukhov.com/PETOUKHOV_ARTICLE_IN_BIOSYSTEMS.pdf.
- Petoukhov S.V.** (2018a). The Genetic Coding System and Unitary Matrices. *Preprints* 2018, 2018040131 (doi: 10.20944/preprints201804.0131.v2).
- Petoukhov S.V.** (2018b). The rules of long DNA-sequences and tetra-groups of oligonucleotides. - [arXiv:1709.04943v5](https://arxiv.org/abs/1709.04943v5), 5th version (8 October 2018), 159 pages.
- Petoukhov S.V.** (2019a). Structural Connections between Long Genetic and Literary Texts. *Preprints* 2018, 2018120142, online 15 02 2019. DOI: 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 2019. *Advances in Intelligent Systems and Computing*, vol 938. Springer, Cham. DOI 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., Petukhova E.S.** (2017). Resonances and the quest for transdisciplinarity. – *Information Studies and the Quest for Transdisciplinarity*. Editors M. Burgin, W.Hofkirchner, World Scientific, p. p. 467-487.
- Petoukhov S.V., Petukhova E.S., 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. P. 588-600. *Advances in Intelligent Systems and Computing*, vol 754. Springer, Cham, DOI: https://doi.org/10.1007/978-3-319-91008-6_59
- Petoukhov S.V., Svirin V.I.** (2012). Fractal genetic nets and the rules of long genetic sequences. – *The Integrated Scientific Journal*, № 8-9, p. 50-52 (in Russian).
- Petoukhov S.V., Svirin V.I.** (2018). The New Wide Class of Symmetries in Long DNA-Texts. Elements of Quantum-Algorithmic Genetics. - *Biologia Serbica*, v. 40, №1, p. 51, Special Edition, ISSN 2334 6590, UDK 57(051). Book of Abstracts, Belgrade Bioinformatics Conference 2018, 18-22 June, 2018, Belgrade, Serbia, page 51.
- Rapoport A.E., Trifonov E.N.** (2012): Compensatory nature of Chargaff's second parity rule. *Journal of Biomolecular Structure and Dynamics*, November, 1-13, DOI:10.1080/07391102.2012.736757
- Rieper E., Anders J., Vedral V.** (2011). Quantum entanglement between the electron clouds of nucleic acids in DNA. - [arXiv:1006.4053v2](https://arxiv.org/abs/1006.4053v2).
- Schrödinger E.** (1952). Are there quantum jumps? Part I. - *The British Journal for the Philosophy of Science*, Vol. 3, No. 10, pp. 109-123. Published by: Oxford University Press on behalf of The British Society for the Philosophy of Science Stable, URL: <http://www.jstor.org/stable/685552>.
- Watson J.D., Crick F.H.C.** (1953). A Structure for Deoxyribose Nucleic Acid. – *Nature*, 171, p. 737-738.