Hyperbolic rules of the oligomer cooperative organization of eukaryotic and prokaryotic genomes

Sergey V. Petoukhov

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 3rd International Conference on Computer Science, Engineering and Education Applications (Kiev, Ukraine, 21-22 January 2020). Also an author's presentation with elements of this article was done at the 6th International Conference in Code Biology (Friedrichsdorf, Germany, 3-7 June 2019, http://www.codebiology.org/conferences/Friedrichsdorf2019/).

Abstract. The author's method of oligomer sums for analysis of oligomer compositions of eukaryotic and prokaryotic genomes is described. The use of this method revealed the existence of general rules for cooperative oligomeric organization of a wide list of genomes. These rules are called hyperbolic because they are associated with hyperbolic sequences including the harmonic progression 1, 1/2, 1/3, ..., 1/n. These rules are demonstrated by examples of quantitative analysis of many genomes from the human genome to the genomes of archaea and bacteria. The hyperbolic (harmonic) rules, speaking about the existence of algebraic invariants in full genomic sequences, are considered as candidates for the role of universal rules for cooperative organization of genomes. The described phenomenological results were obtained as consequences of the previously published author's quantum-information model of long DNA sequences. The oligomer sums method was also applied to the analysis of long genes and viruses including the COVID-19 virus; this revealed, in characteristics of many of them, the phenomenon of rhythmically repeating deviations from model hyperbolic sequences; these deviations are associated with DNA triplets and should be systematically analyzed for a deeper understanding the genetic coding system. The topics of the algebraic harmony in living bodies and of the quantum-information approach in biology are discussed.

Key words. DNA oligomers, harmonic progression, hyperbolic rules, matrices, tensor product, quantum informatics, oligomer sums method, genomes, genes, viruses.

Contents

1. Introduction
2. The hyperbolic rule in the oligomer cooperative organization of all human chromosomes
3. The representation of the DNA alphabets by their binary-oppositional traits in matrix genetics
4. The hyperbolic rules in all chromosomes of a fruit fly Drosophila melanogaster
5. The hyperbolic rules in all chromosomes of a nematode Caenorhabditis elegans
6. The hyperbolic rules in all chromosomes of a house mouse Mus musculus
7. The hyperbolic rules in all chromosomes of a plant Arabidopsis thaliana
Living bodies are huge sets of various molecules, which have an amazing ability to inherit biological traits of organisms to the next generations. G. Mendel, in his experiments with plant hybrids, found that the transmission of traits under the crossing of organisms occurs by certain algebraic rules, despite the colossal heterogeneity and complexity of molecular structures of their bodies. This article represents new results of studying hidden algebraic rules in molecular genetic information structures.

One of the founders of quantum mechanics, who introduced also the term “quantum biology,” P. Jordan noted 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 [McFadden and Al-Khalili, 2018]. Taking into account the dictatorial influence of DNA and RNA molecules on the whole body, the author focused his research on a special analysis of numeric parameters of nucleotide sequences in single-stranded DNA of different genomes and their parts. As a result of this research, a new method of analysis of nucleotide sequences was created, which has led to discovering new numeric rules of cooperative oligomer organization of eukaryotic and prokaryotic genomes. These materials are described below. All initial data on nucleotide sequences for this analysis were taken from the GenBank.

For long nucleotide sequences of single-stranded DNA, the second Chargaff’s rule is well known, which states that in such sequences the amount of guanine G is approximately equal to the amount of cytosine C and the amount of adenine A is approximately equal to the amount of thymine T. Many authors have devoted their works to the analysis and discussion of this rule (see, for example, [Fimmel, Gumbel, Karpuzoglu, Petoukhov, 2019; Prabhu, 1993; Rapoport, Trifonov, 2012; Rosandic, Vlahovic, Gluncic, Paar, 2016; Shporer, Chor, Rosset, and Horn, 2016; Yamagishi, 2017]). 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.

Regarding the quantitative analysis of DNA sequences, researchers usually study quantities and percentages (or probability, or frequencies) of separate \( n \)-plets (that is separate oligomers, having their length \( n \)). For example, the second Chargaff’s rule is based on such a study of the quantities of separate nucleotides A, T, C, and G. The
work [Prahb, 1993] studies quantities of separate n-plets. In contrast to such analytic approaches, the author suggests for analysis of long nucleotide sequences another method called the oligomer sums method. It allows studying oligomer cooperative organization by analysis of total amounts of all n-plets, having fixed length n, from the certain equivalence classes of oligomers. Below this analytic approach and the results of its application to many genomes and separate nucleotide sequences are represented. The presented study is a continuation of long term author's researches on biological symmetries.

2. The hyperbolic rule in the oligomer cooperative organization of all human nuclear chromosomes

The term “oligomer” refers to a molecular complex of chemical that consists of a few repeating units. Nucleobases - adenine A, thymine T, cytosine C, and guanine G - serve as such repeated units in DNA oligomers, which can have different lengths and which are also called n-plets, where n refers to the oligomer length. Each of nucleotide sequences in eukaryotic and prokaryotic genomes can be considered as a sequence of monomers (like as A-C-A-T-G-T-…), or a sequence of doubles (like as AC-AT-GT-GG-…), or a sequence of triplets (like as ACA-TGT-GGA-…), etc. The article describes the numerical analysis of sets of n-plets, which belong to the equivalence classes (or cooperative groupings) of A₁-oligomers, or T₁-oligomers, or C₁-oligomers, or G₁-oligomers correspondingly (their index 1 indicates that all oligomers of each class start with the same nucleotide A, or T, or C, or G). For example, the class of the A₁-oligomers contains the following n-plets: 4 doublets AA, AT, AC, and AG; 16 triplets AAA, AAT, AAC, AAG, ATA, …, AGG; etc. The total amount of different kinds of n-plets, which start with the same nucleotide, under fixed n is equal to 4ⁿ⁻¹.

To simplify a theoretical explanation, let us consider the example of an analysis of the oligomer cooperative organization of human chromosome №1 by the author’s method of oligomer sums (abbreviation, the OS-method). The totality of data obtained by analyzing a nucleotide sequence by the OS-method is called its OS-representations. This method gives numeric sequences called oligomer sums sequences (or briefly, OS-sequences).

The application of the OS-method to the analysis of the human chromosome №1 includes the following steps, which are typical also for the analysis of other DNA and RNA sequences:

- Firstly, one should calculate phenomenological quantities Sₐ, Sₜ, Sₙ, and S₉ of monomers A, T, C, and G correspondingly in the considered nucleotide sequence. In the human chromosome №1, the following quantities exist: Sₐ = 67070277, Sₜ = 67244164, Sₙ = 48055043, S₉ = 48111528;
- Secondly, to construct the oligomer sums sequences, one should calculate the total amounts Σₐₙ, Σₜₙ, Σₙₙ, and Σ₉ₙ of n-plets in equivalence classes of A₁-oligomers, T₁-oligomers, C₁-oligomers, and G₁-oligomers under n = 1, 2, 3, 4, … (here, for example, the symbol Σₐ₃ₙ refers to the total amount of triplets, which start with the nucleotide A). These total amounts regarding each of the classes are members of the appropriate OS-sequences of the class. For analysis of human chromosomes and various eukaryotic and prokaryotic genomes, the author usually takes n = 1, 2, 3, …, 19, 20 or, in special cases, n = 1, 2, 3, …, 99, 100.
One can remind here that genomic sequences in the GenBank sites usually contain some letters N, indicating that there can be any nucleotide in this place (https://www.ncbi.nlm.nih.gov/books/NBK21136/). By this reason, the total amount of all monomers A, T, C, G (that is the sum $S_A + S_T + S_C + S_G$), 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 results of the application of the OS-method to analyze genomic sequences.

For human chromosome №1, phenomenological values of the total amounts of $n$-plets from the class $A_1$-oligomers are shown in the graphical form for $n = 1, 2, 3, \ldots, 20$ in Fig.2.1, left (in blue). Here the abscissa axis represents the values of $n$, and the ordinate axis represents the values of the total amounts $\Sigma_{A,n,1}$ of $n$-plets, which start with the nucleotide A. The amazing result is that all 20 phenomenological points $[n, \Sigma_{A,n,1}]$ lie - with a high level of accuracy - along with the hyperbola $H_{A,1} = S_A/n = 67070277/n$ shown in red in Fig. 2.1, middle. Deviations of phenomenological quantities $\Sigma_{A,n,1}$ from model values $S_A/n$ lie in the range $-0.030\% \div 0.024\%$, that is, they comprise only one-hundredths of a percent (Fig. 2.1, right). Initial data on this chromosome were taken in the GenBank: https://www.ncbi.nlm.nih.gov/nuccore/NC_000001.11.

![Fig. 2.1](image)

**Fig. 2.1.** The graphs of data for the case of the OS-sequences of $n$-plets from the class $A_1$-oligomers of the human chromosome №1. In these graphs, the abscissa axis represents the values of $n$, and the ordinate axis represents the set of phenomenological total amounts $\Sigma_{A,n,1}$ of $n$-plets beginning with the nucleotide A. **Left:** the ordinate axis represents the set of phenomenological total amounts $\Sigma_{A,n,1}$ of $n$-plets beginning with the nucleotide A. **Middle:** the ordinate axis represents modeling values $S_A/n = 67070277/n$. The dots with coordinates $[n, S_A/n]$ belong to the shown hyperbola $H_{A,1} = S_A/n = 67070277/n$. **Right:** deviations of the real OS-sequence $\Sigma_{A,n,1}$ from the model hyperbolic progression $S_A/n$ in percentages.

This result is striking because it shows that knowing only the number of nucleotides A, that is, only one member of the number series shown in Fig. 2.1, left, one can predict with the high accuracy all other 19 members, each of which is a sum of $4^{n-1}$ possible kinds of $n$-plets. The number of possible kinds of $n$-plets in these sums is growing rapidly, becoming astronomically huge: 4, 16, 64, 256, 1024, ..., $4^{10}$, ..., $4^{19}$. Of course, in the human chromosome №1, for example, not all possible $4^{19}$ kinds of the mentioned 20-plets exist but the total amount of all those kinds of 20-plets, which exist in this chromosome, is practically equal to $S_A/20$ with a high level of accuracy shown below.

Similar results were obtained when studying in this chromosome the total amounts of $n$-plets, which start with the nucleotide T (Fig. 2.2, at left), and with the nucleotide C (Fig. 2.2, at middle), and with the nucleotide G (Fig. 2.2, at right). The phenomenological values of the total amounts $\Sigma_{T,n,1}$, $\Sigma_{C,n,1}$, and $\Sigma_{G,n,1}$ of $n$-plets are...
also modeled effectively by appropriate hyperbolic progressions $H_{T,1}$, $H_{C,1}$, $H_{G,1}$ (2.1), which differ from each other only by their numerators $S_T$, $S_C$, and $S_G$:

$$H_{T,1} = \frac{S_T}{n} = \frac{67244164}{n}, \quad H_{C,1} = \frac{S_C}{n} = \frac{48055043}{n}, \quad H_{G,1} = \frac{S_G}{n} = \frac{48111528}{n} \quad (2.1)$$

Fig. 2.2. Additional graph data to the OS-representation of the human chromosome №1. The abscissa axes represent the values $n = 1, 2, 3, \ldots, 20$. The ordinate axes show model values $H_{T,1}(n)$, $H_{C,1}(n)$, and $H_{G,1}(n)$ (in red) from (2.1), which practically coincide phenomenological values $\Sigma_{T,n,1}$, $\Sigma_{C,n,1}$, and $\Sigma_{G,n,1}$ of the total amount of $n$-plets, which start with the nucleotide $T$ (at the left graph), the nucleotide $C$ (at the middle graph), and the nucleotide $G$ (at the right graph). The numerical data on this coincidence is shown below.

Fig. 2.3 shows real and model values for the OS-representation of the classes of $A_1$, $T_1$, $C_1$, and $G_1$-oligomers of the human chromosome №1 for $n = 1, 2, 3, \ldots, 20$. The model values of the total amounts of $n$-plets, which start with a certain nucleotide ($A$, $T$, $C$, or $G$), are calculated correspondingly as values of the hyperbolic progressions $H_{A,1} = \frac{S_A}{n}$, $H_{T,1} = \frac{S_T}{n}$, $H_{C,1} = \frac{S_C}{n}$, and $H_{G,1} = \frac{S_G}{n}$. Deviations of real values from model values are also shown in percent in accordance with the expression: $100(1 - (\text{real value})/\text{(model value)})$. One can see that these deviations are much lesser than 0,2% in all cases.

| $n$ | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 |
|-----|----|----|----|----|----|----|----|----|----|----|
| A   |    |    |    |    |    |    |    |    |    |    |
| Real| 67070277 | 33357519 | 22360413 | 16768845 | 13413532 | 1179286 | 9584038 | 8383461 | 7453552 | 6706672 |
| Model| 67070277 | 33357519 | 22360413 | 16768845 | 13413532 | 1179286 | 9584038 | 8383461 | 7453552 | 6706672 |
| $\Delta$%| 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| T   |    |    |    |    |    |    |    |    |    |    |
| Real| 67244164 | 33620498 | 22421293 | 16808862 | 13445386 | 11712704 | 9606748 | 8405049 | 7470415 | 6724359 |
| Model| 67244164 | 33620498 | 22421293 | 16808862 | 13445386 | 11712704 | 9606748 | 8405049 | 7470415 | 6724359 |
| $\Delta$%| 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| C   |    |    |    |    |    |    |    |    |    |    |
| Real| 48055043 | 24024903 | 16012711 | 12013624 | 9612227 | 8005708 | 6885944 | 6008215 | 5336968 | 4803979 |
| Model| 48055043 | 24024903 | 16012711 | 12013624 | 9612227 | 8005708 | 6885944 | 6008215 | 5336968 | 4803979 |
| $\Delta$%| 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| G   |    |    |    |    |    |    |    |    |    |    |
| Real| 48111528 | 24057608 | 16049889 | 12028924 | 9625086 | 8021235 | 6869132 | 6013412 | 5348337 | 4813156 |
| Model| 48111528 | 24057608 | 16049889 | 12028924 | 9625086 | 8021235 | 6869132 | 6013412 | 5348337 | 4813156 |
| $\Delta$%| 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
Fig. 2.3. Real and model values to the OS-representations of the classes of A₁-, T₁-, C₁-, and G₁-oligomers in human chromosome № 1 are shown for \( n = 1, 2, \ldots, 20 \). The real total amounts of \( n \)-plets, which start with a certain nucleotide (A, T, C, or G), are indicated (in blue) jointly with their model values \( H_{A,1}(n) \), \( H_{T,1}(n) \), \( H_{C,1}(n) \), and \( H_{G,1}(n) \) from (2.1) (in red). The symbol \( \Delta \% \) refers to deviations of real values from model values in percent (the model values are taken as 100%).

The model hyperbolic progressions \( H_{A,1} = S_A/n \), \( H_{T,1} = S_T/n \), \( H_{C,1} = S_C/n \), and \( H_{G,1} = S_G/n \) serve as mathematical standards for the described phenomenological facts. These hyperbolic progressions differ from each other only in the magnitude of numerators in their expressions, and therefore they can be specified by the general expression (2.2):

\[
H_{N,1}(n) = S_N/n,
\]

where \( N \) refers to any of nucleotides A, T, C, or G; \( S_N \) refers to the number of corresponding monomers A, T, C, or G in the analyzed nucleotide sequence. If you know the total quantity \( S_N \) of the monomer \( N \), you can predict - with a high level of accuracy - the total amounts of \( n \)-plets belonging to the class \( N_1 \)-oligomers by using the general expression (2.2). These phenomenological facts testify in favor of the cooperative entity of the nucleotide sequence in the human chromosome № 1.

By the corresponding compression of the ordinate axis in these cartesian coordinate systems (that is by appropriate scaling of numerators \( S_A, S_T, S_C, \) and \( S_G \)), each of these four hyperbolic sequences \( H_{A,1}=S_A/n, \) \( H_{T,1}=S_T/n, \) \( H_{C,1}=S_C/n, \) and \( H_{G,1}=S_G/n \) reduces to the hyperbolic sequence (2.3):

\[
y = 1/n,
\]

which we call the canonical (or reference) hyperbolic sequence of OS-representations (or the canonical OS-sequence) of nucleotide sequences. In mathematics, the sequence (2.4)

\[
1/1, 1/2, 1/3, 1/4, 1/5, \ldots, 1/n
\]

is known long ago as a harmonic progression (or a harmonic sequence) where each term is the harmonic mean of the neighboring terms. For this reason, the revealed
hyperbolic sequences in genomes can be also called genomic harmonic progressions, and, in this mathematical sense, one can talk about the harmonic rules and the harmonious organization of genomes described below. The historically famous name "the harmonic progression" comes from the connection (2.4) with the series of harmonics in music. The sums of the first members of the harmonic progression (2.4) are called harmonic numbers. The rich centuries-old history of the study of harmonic progressions and harmonic series is associated with the names of Pythagoras, Orem (d'Oresme), Leibniz, Newton, Euler, Fourier, Dirichlet, Riemann, and other researchers. The generalization of the harmonic series is known as the Riemann zeta function. Using musical terminology, where the term “timbre” refers to the totality of the set of sound frequencies in a prolonged sound, one can conditionally say that the oligomer sums method represents the analyzed nucleotide sequence as some “oligomer timbre”. The series of harmonic numbers serves as the discrete analog of the continuous function of natural logarithm \( \ln n \) [Graham, Knuth, Parashnik, 1994, p. 276]; this, in particular, connects the harmonic progression (2.4) with the Weber-Fechner logarithmic law, which is the main psychophysical law and dictates informatic peculiarities for all inherited sensory channels - vision, hearing, smell, etc., whose organs (eyes, ears, nose, etc.) very differ each other in appearance. It testifies that genetic and different psychophysical levels of inherited biological informatics are structurally intercorrelated on the algebra-harmonical basis [Petoukhov, 2016, 2020b].

Given the relationship of the harmonic progression (2.4) with the four OS-sequences for the four types of nucleotides A, T, C, and G, genomic sequences can be called tetra-harmonic sequences. Fig. 2.3 shows that the OS-sequences of the total amounts of \( n \)-plets from the classes of \( A_1 \)-oligomers and \( T_1 \)-oligomers differ little from each other. The same is true for the OS-sequences of the total amounts of \( n \)-plets from the classes of \( C_1 \)- and \( G_1 \)-oligomers. This fact is described by the expressions (2.5):

\[
\Sigma_{A,n,1} \approx \Sigma_{T,n,1}, \quad \Sigma_{C,n,1} \approx \Sigma_{G,n,1} \tag{2.5}
\]

In the particular case at \( n = 1 \), expressions (2.4) demonstrate the second Chargaff's rule on the approximate equality between the amounts of nucleotides A and T, as well as C and G in long DNA sequences. Correspondingly the phenomenological fact, described by expressions (2.4), is a certain generalization of the 2nd Chargaff’s rule.

The results presented indicate, at least for the human chromosome \#1, that there exist two general hyperbolic (or harmonic) rules regarding the total amounts of \( n \)-plets, which start with a certain nucleotide A, T, C, or G.

The first hyperbolic rule (about interrelations of oligomers in individual chromosomes):

- For any of classes of \( A_1 \)-, \( T_1 \)-, \( C_1 \)-, or \( G_1 \)-oligomers in individual chromosomes, the total amounts \( \Sigma_{N,n,1}(n) \) of their \( n \)-plets, corresponding different \( n \), are interrelated each other through the general expression \( \Sigma_{N,n,1} \approx S_N/n \) with a high level of accuracy (here \( N \) refers to any of nucleotides A, T, C, or G; \( S_N \) refers to the number of monomers \( N \); \( n = 1, 2, 3, 4, \ldots \) is not too large compared to the full length of the nucleotide sequence). The phenomenological points with coordinates \([n, \Sigma_{N,n,1}]\) practically lie on the hyperbola having points \( H_{N,1} = S_N/n \).
• **The second hyperbolic rule** (about the similarity in the pairs of OS-sequences):

In individual chromosomes, two numeric OS-sequences expressing the total amounts of \(n\)-plets, which start with the nucleotide A and with the nucleotide T, are approximately identical. The same is true for two numeric OS-sequences expressing the total amounts of \(n\)-plets, which start with the nucleotide C and with the nucleotide G (in accordance with the expressions (2.5)). Here \(n = 1, 2, 3, 4, \ldots\) is not too large compared to the full length of the nucleotide sequence.

The obtained results of the hyperbolic (or harmonic) interrelationship of the amounts of \(n\)-plets, belonging to the indicated classes of oligomers, are not trivial. Theoretical counter-examples of artificial nucleotide sequences, which have not such interrelation, can be indicated. For example, for the case of the class of \(A_1\)-oligomers, one can mentally construct a long nucleotide sequence that contains many nucleotides A but does not have two adjacent nucleotides A, that is, does not contain a single AA doublet. Such a sequence does not have the hyperbolic interrelationship between the amounts of the nucleotide A and the total amounts of \(n\)-plets starting with A. It can be added that, in the same human chromosome \(\# 1\), the comparison of amounts of \(n\)-plets, consisting only of nucleotides of the same kind, for example, of the nucleotide A, shows the absence of the hyperbolic relationship between them. Really, in this case the amount of the nucleotide A is equal to 67070277, the amount of the doublet AA - 10952057, the amount of the triplet AAA – 2837038, the amount of the tetraplet AAAA – 856207, and so on without their hyperbolic interrelation.

Let us continue the description of obtained results of the analysis of the human genome, which contains 22 autosomes and 2 sex chromosomes X and Y. These chromosomes are very different from each other in length, molecular weight, gene content, etc. What can be said about the other 23 human chromosomes? Are there hyperbolic rules similar to formulated rules for the human chromosome \(\# 1\)? Yes, the author has got a positive answer to this question. For each of 24 human chromosomes, knowing its quantity \(S_N\) of the monomer N (that is A, T, C, or G) allows you to calculate the total amounts of \(n\)-plets, which start with the oligomer N, with a high level of accuracy by using the general expression (2.2). Here \(n = 1, 2, 3, \ldots\) but not very large in comparison with the length of the DNA sequence. Fig. 2.4 shows general confirmational results of studying all 24 human chromosomes by the OS-method under \(n = 1, 2, 3, \ldots, 20\).

These results demonstrate that both hyperbolic (or harmonic) rules \(\# 1\) and \(\# 2\) hold true for each of the human chromosomes with a high level of accuracy.
Fig. 2.4. Some results of the analysis of all 24 human nuclear chromosomes by the oligomer sums method are represented. For each of the chromosomes, quantities $S_A$, $S_T$, $S_C$, and $S_G$ of monomers A, T, C, and G are shown to define the model hyperbolic progressions (2.2). The columns «Range %» show ranges of deviations of real OS-series of corresponding n-plets ($n = 1, 2, \ldots, 20$) from their appropriate model values $S_A/n$, $S_T/n$, $S_C/n$, and $S_G/n$ in percentages (in each case, an appropriate model value is taken as 100%). The left column shows chromosome numbers.

One can show that the obtained phenomenological data also leads to the third hyperbolic rule related to normalized versions of the OS-sequences $S_A/n$, $S_T/n$, $S_C/n$, and $S_G/n$. Scaling the numerators $S_A$, $S_T$, $S_C$, and $S_G$ by dividing by their total amount $S = S_A + S_T + S_C + S_G$, we obtain the corresponding scaling of all these OS-sequences, which are termed as "normalized OS-sequences" (2.6):

$$S_A/(nS), \ S_T/(nS), \ S_C/(nS), \ S_G/(nS) \quad (2.6)$$
It turns out that the normalized OS-sequences of all human chromosomes are similar to each other with a high level of accuracy as Fig. 2.5 shows regarding the first main members $S_A/S$, $S_T/S$, $S_C/S$, and $S_G/S$ of these hyperbolic sequences.

| Chrom | $S_A/S$ | $S_T/S$ | $S_C/S$ | $S_G/S$ | Harmonic mean |
|-------|---------|---------|---------|---------|---------------|
| 1     | 0.2910  | 0.2918  | 0.2085  | 0.2087  | 0.243         |
| 2     | 0.2984  | 0.2993  | 0.2009  | 0.2014  | 0.241         |
| 3     | 0.3013  | 0.3020  | 0.1980  | 0.1986  | 0.239         |
| 4     | 0.3086  | 0.3089  | 0.1910  | 0.1915  | 0.236         |
| 5     | 0.3018  | 0.3032  | 0.1971  | 0.1979  | 0.239         |
| 6     | 0.3021  | 0.302   | 0.1979  | 0.197   | 0.239         |
| 7     | 0.2960  | 0.2970  | 0.2033  | 0.2037  | 0.241         |
| 8     | 0.2994  | 0.2990  | 0.2008  | 0.2008  | 0.240         |
| 9     | 0.2928  | 0.2926  | 0.2074  | 0.2072  | 0.243         |
| 10    | 0.2917  | 0.2929  | 0.2074  | 0.2080  | 0.243         |
| 11    | 0.2920  | 0.2926  | 0.2074  | 0.2080  | 0.243         |
| 12    | 0.2957  | 0.2966  | 0.2035  | 0.2042  | 0.242         |
| 13    | 0.3069  | 0.3079  | 0.1926  | 0.1926  | 0.237         |
| 14    | 0.2945  | 0.2970  | 0.2040  | 0.2045  | 0.242         |
| 15    | 0.2896  | 0.2901  | 0.2097  | 0.2106  | 0.244         |
| 16    | 0.2758  | 0.2784  | 0.2221  | 0.2237  | 0.247         |
| 17    | 0.2730  | 0.2738  | 0.2258  | 0.2273  | 0.248         |
| 18    | 0.3011  | 0.3014  | 0.1987  | 0.1989  | 0.240         |
| 19    | 0.2591  | 0.2615  | 0.2388  | 0.2406  | 0.250         |
| 20    | 0.2778  | 0.2810  | 0.2201  | 0.2211  | 0.247         |
| 21    | 0.2964  | 0.2946  | 0.2047  | 0.2043  | 0.242         |
| 22    | 0.2651  | 0.2648  | 0.2339  | 0.2361  | 0.249         |
| X     | 0.3019  | 0.3029  | 0.1971  | 0.1982  | 0.239         |
| Y     | 0.2985  | 0.3012  | 0.2001  | 0.2001  | 0.240         |

**Fig. 2.5.** Data for normalized OS-sequences $S_A/(nS)$, $S_T/(nS)$, $S_C/(nS)$, and $S_G/(nS)$ of all human chromosomes are shown for comparison. Here $S = S_A + S_T + S_C + S_G$. Harmonic means of the values $S_A/S$, $S_T/S$, $S_C/S$, and $S_G/S$ in each chromosome are also indicated.

The same results on the similarity of normalized OS-sequences $S_A/nS$, $S_T/nS$, $S_C/nS$, and $S_G/nS$ in all chromosomes of a particular genome were obtained by the author when studying the genomes of a number of eukaryotes (until now, without a single exception in analyzed cases). Below Sections 4-7 present appropriate results for some eukaryotic genomes. These results allow proposing the third hyperbolic (or harmonic) rule on the total amounts of $n$-plets, which start with a certain nucleotide A, T, C, or G.

**The third hyperbolic rule** (about the similarity of chromosomes):

- All chromosomes of any individual eukaryotic genome have approximately the same normalized OS-sequences $S_A/nS$, $S_T/nS$, $S_C/nS$, and $S_G/nS$ representing classes of $A_1$, $T_1$, $C_1$, and $G_1$-oligomers ($n = 1, 2, 3, 4, \ldots$ is not too large compared to the full length of the nucleotide sequence).
The author suggests that these hyperbolic rules are universal genetic rules. But at this stage of the study, they are only candidates for the role of universal rules, since the analysis of the widest variety of genomes is required to verify their universality.

Let us return to the harmonic progression (2.4) and recall its relation with the well-known concept of the harmonic mean. The harmonic mean $H$ of the positive real numbers $x_1, x_2, ..., x_n$, is defined to be

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}}$$  

(2.7)

Knowing two neighboring members of the harmonic progression, one can calculate its next member. Since the harmonic mean is related to harmonic progressions, the author indicates values of the harmonic mean in some figures of the article for comparison analysis of OS-sequences in different nucleotide sequences (Fig. 2.5 and many others).

Each genomic DNA sequence with its total amount $S$ of all nucleotides A, T, C, and G also contains total amounts $S/n$ of $n$-plets (that is, $S/2$ doublets, $S/3$ triplets, etc.). These total amounts correspond to the hyperbolic sequence $S$, $S/n$, $S/2$, $S/3$, ..., $S/n$. Each member of this sequence is the sum of the four OS-sequences $S_A/n$, $S_T/n$, $S_C/n$, and $S_G/n$ (2.8):

$$S_A/n + S_T/n + S_C/n + S_G/n = S/n \quad \text{or} \quad S_A/nS + S_T/nS + S_C/nS + S_G/nS = 1/n \quad (2.8)$$

These linear superpositions are valid for a wide variety of genomes that differ only in individual coefficients $S_A$, $S_T$, $S_C$, and $S_G$.

Below Sections 4-7 represents the results, which have been obtained on the basis of the analysis of very different genomes by the OS-method and which testify in favor that the formulated hyperbolic rules have a general genomic significance. But previously the next Section will explain the matrix-algebraic approach, which has led the author to discover these rules.

3. The representation of the DNA alphabets by their binary-oppositional traits in matrix genetics

Science does not know why the DNA alphabet of nucleotides consists of only 4 relatively simple molecules A, T, C, and G. But science knows that this alphabet is endowed with a system of binary-opposition traits (or indicators):

- 1) in the double helix of DNA, there are two complementary pairs of nucleotides:
  - the nucleotides C and G of the first pair are connected by three hydrogen bonds, and the nucleotides A and T of the second pair by two hydrogen bonds. Given these oppositional indicators, one can represent $C = G = 1$ and $A = T = 0$;
  - 2) the two nucleotides are keto molecules (G and T), and the other two are amino molecules (A and C). Given these oppositional indicators, one can represent $G = T = 1$ and $A = C = 0$.

Taking this into account, it is convenient to represent DNA alphabets of 4 nucleotides, 16 doublets and 64 triplets in the form of square tables, the columns of which are numbered in accordance with oppositional indicators “3 or 2 hydrogen
bonds” \((C = G = 1, A = T = 0)\), and the rows in accordance with oppositional indicators “amino or keto” \((C = A = 1, G = T = 0)\). In such tables, all letters, doublets, and triplets automatically occupy their strictly individual places (Fig. 3.1).

![Table](image)

**Fig. 3.1.** The square tables of DNA-alphabets of 4 nucleotides, 16 doublets, and 64 triplets with a strict arrangement of all components. Each of the tables is constructed in line with the principle of binary numeration of its column and rows on the basis of binary-oppositional indicators of nucleobases G, T, C, and A (see explanations in the text).

These three tables (Fig. 3.1) are not only simple tables but they are members of the tensor family of matrices: the second and the third tensor (Kronecker) powers of the matrix \([G, T; C, A]\) generate similar arrangements of 16 doublets and 64 triplets inside matrices \([G, T; C, A]^{(2)}\) and \([G, T; C, A]^{(3)}\) as shown in Fig. 3.1. One can note here that the classes of \(G_1\)-, \(T_1\)-, \(C_1\)-, and \(A_1\)-oligomers, analyzed in the previous Section as related to the hyperbolic rules, are connected by a special manner with the tensor family of the matrices \([G, T; C, A]^{(n)}\) where the symbol \((n)\) refers to an appropriate tensor power. More precisely, in Fig. 3.1, each of \((2*2)\)-quadrants of the matrix \([G, T; C, A]^{(2)}\) contains a complete set of 4 doublets, which start with one of nucleotides G, T, C, and A; each of \((2*2^2)\)-quadrants of the matrix \([G, T; C, A]^{(3)}\) contains a complete set of 16 triplets, which start with one of the nucleotides G, T, C, and A. In general, each of \((2^{n-1}*2^{n-1})\)-quadrants of the matrix \([G, T; C, A]^{(n)}\) contains a complete set of \(4^{n-1}\ n\)-plets, which start with one of the nucleotides G, T, C, and A.

The genetic code is called a "degenerate code" because 64 triplets encode 20 amino acids and stop-codons so that several triplets can encode each amino acid at once, and each triplet necessarily encodes only a single amino acid or a stop-codon. The \((8*8)\)-matrix of 64 triplets (Fig. 3.1) was built formally without any mention of amino acids and stop-codons. Nothing data preliminary exist on a possible correspondence between triplets and amino acids. How can these 20 amino acids and stop-codons be located in this matrix of 64 triplets? There are a huge number of
possible options for the location and repetition of separate amino acids and stop-codons in 64 cells of this matrix. More precisely, the number of these options is much more than $10^{100}$ (for comparison, the entire time of the Universe existence is estimated in modern physics at $10^{17}$ seconds). But Nature uses - from this huge number of options - only a very specific repetition and arrangement of separate amino acids and stop-codons, the analysis of which is important for revealing the structural organization of the information foundations of living matter.

Fig. 3.2 shows the real repetition and location of amino acids and stop-codons in the Vertebrate Mitochondrial Code, which is the most symmetrical among known dialects on the genetic code. This genetic code is called the most ancient and "ideal" in genetics [Frank-Kamenetskii, 1988] (other dialects of the genetic code have small differences from this one, which is considered in the theory of symmetries as the basis from the structural point of view).

|     | 111 | 110 | 101 | 100 | 011 | 010 | 001 | 000 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 111 | PRO | PRO | HIS | GLN | THR | THR | ASN | LYS |
| 110 | PRO | PRO | GLN | HIS | THR | THR | LYS | ASN |
| 101 | ARG | ARG | LEU | LEU | SER | STOP | ILE | MET |
| 100 | ARG | ARG | LEU | STOP | SER | MET | ILE | ATT |
| 011 | ALA | ALA | ASP | GLU | SER | SER | TYR | STOP |
| 010 | ALA | ALA | GLU | ASP | SER | SER | STOP | TYR |
| 001 | GLY | GLY | VAL | TRP | CYS | PHE | LEU | TTA |
| 000 | GLY | GLY | VAL | TRP | CYS | LEU | PHE | TTA |

Fig. 3.2. The location and repetition of 20 amino acids and 4 stop-codons (denoted by bold) in the matrix of 64 triplets $[C, A; G, T]_3$ (Fig. 3.1) for the Vertebrate Mitochondrial Code. The symbol “Stop” refers to stop-codons.

The location and repetition of all amino acids and stop-codons in the matrix of 64 triplets have the following algebraic feature (Fig. 3.2):

- Each of sixteen $(2*2)$-sub-quadrants, forming this genetic matrix and denoted by bold frames, is bisymmetrical: each of its both diagonals contains an identical kind of amino acids or stop-codon.

Bisymmetric $(2*2)$-matrices $[a, b; b, a]$ are well known in algebra as matrix representations of two-dimensional hypercomplex numbers called hyperbolic numbers: $a+bj$ where “$a$” and “$b$” are real numbers, and the imaginary unit $j$ satisfies $j^2 = +1$. [Kantor, Solodovnikov, 1989]. Hyperbolic numbers are used in physics and mathematics and they have also synonymical names: "split-complex numbers", "double numbers" and "perplex numbers". The collection of all hyperbolic numbers forms algebra over the field of real numbers [Harkin, Harkin, 2004; Kantor, Solodovnikov, 1989]. The algebra is not a division algebra or field since it contains
zero divisors. Addition and multiplication of hyperbolic numbers are defined by the expressions (3.1):

\[(x+jy)+(u+jv)=(x+u)+j(y+v); \quad (x+jy)(u+jv)=(xu+yv)+j(xv+yu)\]  

(3.1)

This multiplication is commutative, associative, and distributes over addition.

Hyperbolic numbers have the matrix form of their representation in a form of bisymmetric matrix \([a, b; b, a]\). Fig. 3.3 shows the decomposition of such matrix into two sparse matrices, the first of which is the matrix representation of the real unit and the second one is the matrix representation of the imaginary unit \(j\).

![Fig. 3.3](image)

**Fig. 3.3.** The decomposition of the bisymmetric matrix \([a, b; b, a]\) into two sparse matrices representing real and imaginary units of hyperbolic numbers correspondingly.

Regarding the hyperbolas from the hyperbolic rules of the previous Section (Figs. 2.1, 2.2, etc.), it can be noted that the transformation of one point of the hyperbola to another point is determined by the transformation of the hyperbolic rotation, in which the hyperbole glides along with itself. Such a transformation is determined by a bisymmetric matrix \([a, b; b, a]\) representing a special form of hyperbolic numbers (the hyperbolic rotation is known in the special theory of relativity under the name of the Lorentz transformation).

If each amino acid and stop-codon is represented by some characteristic parameter (for example, the number of carbon atoms in these organic formations or numbers of protons in its molecular structure, etc.), then a numerical \((8*8)\)-matrix arises (Fig. 3.4) with bisymmetric \((2*2)\)-sub-quadrants representing hyperbolic numbers \(a+bj\). In other words, this phenomenologic arrangement of amino acids and stop-codons in the matrix of 64 triplets is associated to the multiblock union of matrix presentations of 16 two-dimensional hyperbolic numbers.

![Fig. 3.4](image)

**Fig. 3.4.** The numeric analog of the symbolic \((8*8)\)-matrix of amino acids and stop-codons from Fig. 3.2 for the case of representing each of amino acids by numbers of its carbon atoms (stop-codons are conditionally represented by zero).

The connection of the genetic code with hyperbolic numbers supplements the following statement of the author, presented in a number of his publications.
shown cases, the deviations of real sequences from the G
biology data tables for such large chromosomes of a few model eukaryotic organisms,
method of matrix analysis and modeling combined under the general name "matrix genetics" [Petoukhov, 2008, 2011, 2016, 2017, 2019b,c; Petoukhov, He, 2010; Petoukhov, Petukhova, 2017a,b].

Let's continue the presentation of confirmational data on the existence of hyperbolic (or harmonic) rules in the cooperative oligomeric organization of the eukaryotic and prokaryotic genomes.

4. The hyperbolic rules in all chromosomes of a fruit fly Drosophila melanogaster

This and upcoming Sections 5-7 are devoted to the analysis - by the oligomer sums method (the OS-method) - of single-stranded DNA sequences of the complete sets of chromosomes of a few model eukaryotic organisms, which are used long ago in the study of genetics, development, and disease. Represented tabular data confirm that both hyperbolic (harmonic) rules regarding n-plets from the classes of A1-, T1-, C1-, and G1-oligomers hold for each of described chromosomes at \( n = 1, 2, 3, 4, \ldots, 19, 20 \) (although these rules are also satisfied for larger values of \( n \), at least up to \( n = 100 \), but the data tables for such large \( n \) are too cumbersome to include in the article).

Let us start with a fruit fly Drosophila melanogaster, which is studied in biology labs for over eighty years. All initial data about its chromosomes were taken from the GenBank - https://www.ncbi.nlm.nih.gov/genome/?term=drosophila+melanogaster. Resulting data in Fig. 4.1 confirm that - for all the chromosomes - the model hyperbolic progressions \( H_{A,1}(n) = S_A/n, H_{T,1}(n) = S_T/n, H_{C,1}(n) = S_C/n, \) and \( H_{G,1}(n) = S_G/n \) from the expression (2.2) practically coincide with the real sequences of total amounts of n-plets from the classes A1-, T1-, C1-, and G1-oligomers at \( n = 1, 2, 3, \ldots, 20 \). In all shown cases, the deviations of real sequences from model hyperbolic progressions are
less than 1% as data in the tabular columns «Range %» indicates. This means that the formulated hyperbolic (harmonic) rules are fulfilled in the considered genome.

| №  | S_A | Range % | S_T | Range % | S_C | Range % | S_G | Range % |
|----|-----|---------|-----|---------|-----|---------|-----|---------|
| X  | 6732793 | -0.196 | 6774766 | -0.125 | 4975870 | -0.198 | 4992722 | -0.148 |
| 2L | 6853032 | -0.217 | 6836080 | -0.219 | 4912017 | -0.239 | 4912383 | -0.251 |
| 2R | 7272860 | -0.259 | 7235562 | -0.144 | 5395216 | -0.195 | 5376598 | -0.222 |
| 3L | 8143548 | -0.142 | 8198331 | -0.126 | 5825673 | -0.211 | 5824515 | -0.262 |
| 3R | 9205526 | -0.143 | 9197619 | -0.145 | 6833716 | -0.170 | 6817898 | -0.231 |
| 4  | 425241  | -1.759 | 436669  | -0.423 | 232566  | -1.463 | 236655  | -0.855 |
| Y  | 1056780 | -0.494 | 1008635 | -0.125 | 682725  | -0.268 | 661579  | -0.512 |

**Fig. 4.1.** The results of the analysis of all chromosomes of *Drosophila melanogaster* by the OS-method. The left column shows symbols of chromosomes. S_A, S_T, S_C, and S_G refer to the quantities of monomers A, T, C, and G in appropriate chromosomes. The columns “Range %” show deviations of real sequences from the model hyperbolic progressions H_{A,1}(n) = S_A/n, H_{T,1}(n) = S_T/n, H_{C,1}(n) = S_C/n, and H_{G,1}(n) = S_G/n at n = 1, 2, 3, ..., 20 (the model values are taken as 100%).

**Fig. 4.2** shows data of normalized OS-sequences for all chromosomes of *Drosophila melanogaster*.

| Chrom | S_A/S | S_T/S | S_C/S | S_G/S | Harmonic mean |
|-------|-------|-------|-------|-------|---------------|
| X     | 0.2868| 0.2886| 0.2120| 0.2127| 0.244         |
| 2L    | 0.2915| 0.2907| 0.2089| 0.2089| 0.243         |
| 2R    | 0.2877| 0.2862| 0.2134| 0.2127| 0.245         |
| 3L    | 0.2909| 0.2929| 0.2081| 0.2081| 0.243         |
| 3R    | 0.2872| 0.2869| 0.2132| 0.2127| 0.245         |
| 4     | 0.3195| 0.3280| 0.1747| 0.1778| 0.228         |
| Y     | 0.3099| 0.2958| 0.2002| 0.1940| 0.239         |

**Fig. 4.2.** Data of normalized OS-sequences S_A/nS, S_T/nS, S_C/nS, and S_G/nS of all chromosomes of *Drosophila melanogaster* are shown for comparison. Here S = S_A+S_T+S_C+S_G. Harmonic means of the values S_A/S, S_T/S, S_C/S, and S_G/S in each chromosome are also indicated.

**5. The hyperbolic rules in all chromosomes of a nematode *Caenorhabditis elegans***

The Section represents data about results of the analysis of single-stranded DNA sequences of the complete set of chromosomes of free-living soil nematode *Caenorhabditis elegans* by the OS-method. This nematode is widely used as a model organism in genetics for a long time. The *Caenorhabditis elegans* nuclear genome is approximately 100 Mb, distributed among six chromosomes. All initial data are taken.
from the GenBank (https://www.ncbi.nlm.nih.gov/genome?term=caenorhabditis%20elegans).

Resulting data in Fig. 5.1 confirm that - for all the chromosomes - the model hyperbolic progressions \( H_{A,i}(n) = S_A/n \), \( H_{T,i}(n) = S_T/n \), \( H_{C,i}(n) = S_C/n \), and \( H_{G,i}(n) = S_G/n \) from the expression (2.2) practically coincide with the real sequences of total amounts of \( n \)-plets from the classes \( A_1 \), \( T_1 \), \( C_1 \), and \( G_1 \)-oligomers at \( n = 1, 2, 3, \ldots, 20 \). In all shown cases, the deviations of real sequences from model hyperbolic (harmonic) progressions are less than 0.5% as data in the tabular columns «Range %» indicates. This means that the formulated hyperbolic (harmonic) rules are fulfilled in the considered genome.

| №  | \( S_A \) | Range % | \( S_T \) | Range % | \( S_C \) | Range % | \( S_G \) | Range % |
|-----|-----------|--------|-----------|--------|-----------|--------|-----------|--------|
| 1   | 4835939   | -0.144 | 4848450   | -0.160 | 2695890   | -0.487 | 2692155   | -0.498 |
| 2   | 4878209   | -0.196 | 4869734   | -0.229 | 2769232   | -0.256 | 2762246   | -0.253 |
| 3   | 4444681   | -0.139 | 4423618   | -0.269 | 2449158   | -0.451 | 2466344   | -0.173 |
| 4   | 5711043   | -0.106 | 5730974   | -0.253 | 3034784   | -0.393 | 3017028   | -0.199 |
| 5   | 6750403   | -0.145 | 6760297   | -0.164 | 3712075   | -0.222 | 3701405   | -0.418 |
| X   | 5747200   | -0.256 | 5734092   | -0.166 | 3119741   | -0.156 | 3117909   | -0.272 |

**Fig. 5.1.** The results of the analysis of all chromosomes of *Caenorhabditis elegans* by the OS-method. The left column shows symbols of chromosomes. \( S_A \), \( S_T \), \( S_C \), and \( S_G \) refer to the quantities of monomers A, T, C, and G in appropriate chromosomes. The columns “Range %” show deviations of real series from the model hyperbolic progressions \( H_{A,i}(n) = S_A/n \), \( H_{T,i}(n) = S_T/n \), \( H_{C,i}(n) = S_C/n \), and \( H_{G,i}(n) = S_G/n \) at \( n = 1, 2, 3, \ldots, 20 \) (the model values are taken as 100%).

**Fig. 5.2.** Data of normalized OS-sequences for all chromosomes of *Caenorhabditis elegans.*

| Chrom | \( S_A/S \) | \( S_T/S \) | \( S_C/S \) | \( S_G/S \) | Harmonic mean |
|-------|-------------|-------------|-------------|-------------|--------------|
| 1     | 0.3208      | 0.3217      | 0.1789      | 0.1786      | 0.230        |
| 2     | 0.3193      | 0.3187      | 0.1812      | 0.1808      | 0.231        |
| 3     | 0.3225      | 0.3209      | 0.1777      | 0.1789      | 0.229        |
| 4     | 0.3265      | 0.3276      | 0.1735      | 0.1725      | 0.226        |
| 5     | 0.3226      | 0.3231      | 0.1774      | 0.1769      | 0.229        |
| X     | 0.3244      | 0.3236      | 0.1761      | 0.1760      | 0.228        |

**Fig. 5.2.** Data of normalized OS-sequences \( S_A/nS, \ S_T/nS, \ S_C/nS, \) and \( S_G/nS \) of all chromosomes of *Caenorhabditis elegans* are shown for comparison. Here \( S = S_A + S_T + S_C + S_G \). Harmonic means of the values \( S_A/S, \ S_T/S, \ S_C/S, \) and \( S_G/S \) in each chromosome are also indicated.
6. The hyperbolic rules in all chromosomes of a house mouse *Mus musculus*

The Section represents data about results of the analysis of single-stranded DNA sequences of the complete set of chromosomes of the laboratory mouse *Mus musculus*, which is a major model organism for basic mammalian biology, human disease, and genome evolution. All initial data are taken from the GenBank https://www.ncbi.nlm.nih.gov/genome?term=mus%20musculus.

Resulting data in Fig. 6.1 confirm that - for all the chromosomes - the model harmonic progressions $H_{A,n}(n) = S_A/n$, $H_{T,n}(n) = S_T/n$, $H_{C,n}(n) = S_C/n$, and $H_{G,n}(n) = S_N/n$ from the expression (2.2) practically coincide with the real sequences of total amounts of $n$-plets from the classes $A_1$, $T_1$, $C_1$, and $G_1$-oligomers at $n = 1$, 2, 3, …, 20. In all shown cases, the deviations of real sequences from model hyperbolic progressions are significantly less than 0.5% as data in the tabular columns «Range %» indicates. This means that the formulated hyperbolic (harmonic) rules are fulfilled in the considered genome.

| №  | $S_A$  | Range % | $S_T$  | Range % | $S_C$  | Range % | $S_G$  | Range % |
|----|--------|---------|--------|---------|--------|---------|--------|---------|
| 1  | 56530182 | -0.051  | 56416289 | -0.126  | 39495313 | -0.044  | 39467408 | -0.150  |
| 2  | 51600126 | -0.099  | 51679955 | -0.060  | 37504114 | -0.041  | 37542456 | -0.110  |
| 3  | 46503996 | -0.041  | 46631177 | -0.057  | 31603703 | -0.151  | 31659979 | -0.063  |
| 4  | 43821952 | -0.050  | 43922197 | -0.056  | 32146231 | -0.097  | 32165231 | -0.050  |
| 5  | 42488105 | -0.060  | 42515761 | -0.063  | 31456650 | -0.033  | 31459158 | -0.218  |
| 6  | 42843713 | -0.122  | 42886213 | -0.107  | 30315703 | -0.028  | 30290914 | -0.083  |
| 7  | 40271749 | -0.105  | 40595947 | -0.110  | 30554235 | -0.089  | 30519876 | -0.019  |
| 8  | 36224525 | -0.043  | 36167473 | -0.119  | 26616967 | -0.147  | 26602467 | -0.123  |
| 9  | 34722476 | -0.079  | 34694585 | -0.088  | 25880876 | -0.154  | 25859081 | -0.114  |
| 10 | 37185184 | -0.066  | 37273294 | -0.046  | 26277876 | -0.099  | 26331308 | -0.112  |
| 11 | 33401283 | -0.069  | 33317397 | -0.121  | 26022668 | -0.057  | 26004597 | -0.061  |
| 12 | 33897029 | -0.111  | 34225639 | -0.105  | 24374340 | -0.061+0.094 | 24425412 | -0.115  |
| 13 | 34255191 | -0.119  | 34115119 | -0.064  | 24377641 | -0.079+0.119 | 24373242 | -0.138  |
| 14 | 35695406 | -0.135  | 35758968 | -0.086  | 24980458 | -0.057+0.135 | 25007278 | -0.099  |
| 15 | 29177752 | -0.135  | 29244798 | -0.028  | 21121081 | -0.050+0.180 | 21109684 | -0.095  |
| 16 | 28035438 | -0.054  | 28084677 | -0.067  | 19439086 | -0.125+0.023 | 19460557 | -0.170  |
| 17 | 26251160 | -0.075  | 26302830 | -0.051  | 19586991 | -0.085+0.060 | 19566481 | -0.096  |
Fig. 6.1. The results of the analysis of all chromosomes of a house mouse *Mus musculus* by the OS-method. The left column shows symbols of chromosomes. $S_A$, $S_T$, $S_C$, and $S_G$ refer to the quantities of monomers A, T, C, and G in appropriate chromosomes. The columns “Range %” show deviations of real sequences from the model hyperbolic sequences $H_A(1) = S_A/n$, $H_T(1) = S_T/n$, $H_C(1) = S_C/n$, and $H_G(1) = S_G/n$ at $n = 1, 2, 3, \ldots, 20$ (the model values are taken as 100%).

Fig. 6.2 shows data of normalized OS-sequences for all chromosomes of *Mus musculus*.

| Chrom | $S_A/S$ | $S_T/S$ | $S_C/S$ | $S_G/S$ | Harmonic mean |
|-------|---------|---------|---------|---------|---------------|
| 1     | 0.2946  | 0.2940  | 0.2058  | 0.2057  | 0.242         |
| 2     | 0.2894  | 0.2898  | 0.2103  | 0.2105  | 0.244         |
| 3     | 0.2973  | 0.2982  | 0.2021  | 0.2024  | 0.241         |
| 4     | 0.2882  | 0.2889  | 0.2114  | 0.2115  | 0.244         |
| 5     | 0.2872  | 0.2874  | 0.2127  | 0.2127  | 0.244         |
| 6     | 0.2928  | 0.2931  | 0.2072  | 0.2070  | 0.243         |
| 7     | 0.2839  | 0.2856  | 0.2154  | 0.2151  | 0.245         |
| 8     | 0.2884  | 0.2879  | 0.2119  | 0.2118  | 0.244         |
| 9     | 0.2866  | 0.2864  | 0.2136  | 0.2134  | 0.245         |
| 10    | 0.2926  | 0.2933  | 0.2068  | 0.2072  | 0.243         |
| 11    | 0.2813  | 0.2806  | 0.2191  | 0.2190  | 0.246         |
| 12    | 0.2899  | 0.2927  | 0.2085  | 0.2089  | 0.243         |
| 13    | 0.2925  | 0.2913  | 0.2081  | 0.2081  | 0.243         |
| 14    | 0.2939  | 0.2945  | 0.2057  | 0.2059  | 0.242         |
| 15    | 0.2899  | 0.2905  | 0.2098  | 0.2097  | 0.244         |
| 16    | 0.2950  | 0.2956  | 0.2046  | 0.2048  | 0.242         |
| 17    | 0.2862  | 0.2868  | 0.2136  | 0.2134  | 0.245         |
| 18    | 0.2929  | 0.2927  | 0.2069  | 0.2075  | 0.243         |
| 19    | 0.2875  | 0.2852  | 0.2139  | 0.2134  | 0.245         |
| X     | 0.3038  | 0.3037  | 0.1962  | 0.1963  | 0.239         |
| Y     | 0.3046  | 0.3065  | 0.1949  | 0.1940  | 0.238         |

Fig. 6.2. Data of normalized OS-sequences $S_A/nS$, $S_T/nS$, $S_C/nS$, and $S_G/nS$ of all chromosomes of a house mouse *Mus musculus* are shown for comparison. Here $S = S_A + S_T + S_C + S_G$. Harmonic means of the values $S_A/S$, $S_T/S$, $S_C/S$, and $S_G/S$ in each chromosome are also indicated.
7. The hyperbolic rules in all chromosomes of a plant *Arabidopsis thaliana*

One more model organism is a plant *Arabidopsis thaliana*. This small flowering plant is used for over fifty years to study plant mutations and for classical genetic analysis. It became the first plant genome to be fully sequenced; it has a small genome of ~120 Mb. The Section represents data about results of the analysis of single-stranded DNA sequences of the complete set of 5 chromosomes of this plant by the oligomer sums method. All initial data about the chromosomes were taken from the GenBank - [https://www.ncbi.nlm.nih.gov/genome/4](https://www.ncbi.nlm.nih.gov/genome/4), the column RefSeq).

Resulting data in Fig. 7.1 confirm that - for all the chromosomes - the model harmonic progressions $H_{A,1}(n) = S_A/n$, $H_{T,1}(n) = S_T/n$, $H_{C,1}(n) = S_C/n$, and $H_{G,1}(n) = S_G/n$ from the expression (2.2) practically coincide with the real sequences of total amounts of $n$-plets from the classes $A_1$, $T_1$, $C_1$, and $G_1$-oligomers at $n = 1, 2, 3, \ldots, 20$. In all shown cases, the deviations of real sequences from model harmonic progressions are less than 0.6% as data in the tabular columns «Range %» indicates. This means that the formulated hyperbolic (harmonic) rules are fulfilled in the considered genome.

| № | $S_A$ | Range % | $S_T$ | Range % | $S_C$ | Range % | $S_G$ | Range % |
|---|-------|---------|-------|---------|-------|---------|-------|---------|
| 1 | 9709674 | -0.275 | 9697113 | -0.140 | 5435374 | -0.130 | 5421151 | -0.186 |
| 2 | 6315641 | -0.035 | 6316348 | -0.256 | 3542973 | -0.406 | 3520766 | -0.148 |
| 3 | 7484757 | -0.121 | 7448059 | -0.283 | 4258333 | -0.141 | 4262704 | -0.206 |
| 4 | 5940546 | -0.155 | 5914038 | -0.238 | 3371349 | -0.222 | 3356091 | -0.206 |
| 5 | 8621974 | -0.123 | 8652238 | -0.213 | 4832253 | -0.515 | 4858759 | -0.132 |

Fig. 7.1. The results of the analysis of all chromosomes of a plant *Arabidopsis thaliana* by the OS-method. The left column shows symbols of chromosomes. $S_A$, $S_T$, $S_C$, and $S_G$ refer to the quantities of monomers A, T, C, and G in appropriate chromosomes. The columns “Range %” show deviations of real sequences from the model hyperbolic progressions $H_{A,1}(n) = S_A/n$, $H_{T,1}(n) = S_T/n$, $H_{C,1}(n) = S_C/n$, and $H_{G,1}(n) = S_G/n$ at $n = 1, 2, 3, \ldots, 20$ (the model values are taken as 100%).

Fig. 7.2 shows data of normalized OS-sequences for all chromosomes of *Arabidopsis thaliana*.

| Chrom | $S_A/S$ | $S_T/S$ | $S_C/S$ | $S_G/S$ | Harmonic mean |
|-------|---------|---------|---------|---------|---------------|
| 1     | 0.3208  | 0.3204  | 0.1796  | 0.1791  | 0.230         |
| 2     | 0.3207  | 0.3207  | 0.1799  | 0.1788  | 0.230         |
| 3     | 0.3191  | 0.3176  | 0.1816  | 0.1817  | 0.231         |
| 4     | 0.3197  | 0.3183  | 0.1814  | 0.1806  | 0.231         |
| 5     | 0.3197  | 0.3209  | 0.1792  | 0.1802  | 0.230         |

Fig. 7.2. Data of normalized OS-sequences $S_A/nS$, $S_T/nS$, $S_C/nS$, and $S_G/nS$ of all chromosomes of *Arabidopsis thaliana* are shown for comparison. Here
S = S_A + S_T + S_C + S_G. Harmonic means of the values S_A/S, S_T/S, S_C/S, and S_G/S in each chromosome are also indicated.

8. Analysis of long genes by the oligomer sums method

Before proceeding to the analysis of prokaryotic genomes, it is useful to show the applicability of the oligomer sum method to the analysis of genes whose sequences are much shorter than DNA sequences in chromosomes. The application of the method unexpectedly reveals the phenomenon of regular rhythmic deviations of the sequences of real total sums of n-plets in the described genes from the corresponding model hyperbolic progressions.

Let us first consider the human TTN gene encoding the largest known protein Titin. Titin, also known as connectin, is important in the contraction of striated muscle tissues. Figs. 8.1–8.6 show some results of the analysis - by the oligomer sums method - of the nucleotide sequence of the TTN gene (numeric results will be represented below). Initial data on its nucleotide sequence are taken in the GenBank https://www.ncbi.nlm.nih.gov/nuccore/X90568.1. This gene contains 26373 nucleotides A, 19569 nucleotides T, 17097 nucleotides C, and 18901 nucleotides G, that is S_A = 26373, S_T = 19569, S_C = 17097, and S_G = 18901 for the model hyperbolic progressions (2.2). It can be especially noted that, in this gene, the amounts of nucleotides A and T are significantly different (26373 and 19569), that is, the second Chargaff’s rule on their approximate equality in long sequences is not satisfied here since this nucleotide sequence is not enough long for the Chargaff’s rule.

The class of T_1-oligomers The class of G_1-oligomers

Fig. 8.1. Graphical representations of the results of the analysis - by the oligomer sums method – of the human TTN gene. The OS-sequences of its total amounts of n-plets, which start with the nucleotide T (left) and the nucleotide G (right), are shown. The red lines refer to model hyperbolic progressions S_T/n and S_G/n correspondingly, where S_T = 19569 and S_G = 18901 are quantities of nucleotides T and G in the gene; n = 1, 2, 3, ..., 20 as shown at the abscissa axes. The blue line (left) and the green line (right) with dots on them refer to the real OS-sequences of the total amounts of such n-plets. The ordinate axes indicate the total amounts of n-plets.

Fig. 8.1 shows the sequences of the highly regular significant deviations of the real total amounts of n-plets, which start with the nucleotide T and the nucleotide G, from model hyperbolic progressions S_T/n = 19569/n and S_G/n = 18901/n. One should note that all these significant deviations happen only at n = 3, 6, 9, ..., 3m, that is only for cases of 3m-plets (here m = 1, 2, 3,...). Correspondingly these significant deviations can be called «triplet-deviations».
Fig. 8.2 shows the graph, which unites both graphs from Fig. 8.1 and demonstrates a few interesting features of the highly regular series of these triplet-deviations.

![Graph](image)

**Fig. 8.2.** The graph, uniting two graphs from Fig. 8.1 for the TNT gene, is shown. The blue dot line and the green dot lines correspond to those additional hyperbolic progressions 11979/n and 28788/n, which model real total amounts of 3m-plets. Other parts of this united graph are the same as in Fig. 8.1.

Firstly, one can see in Fig. 8.2 that, in classes of T1-oligomers and G1-oligomers, the triplet-deviations happen in opposite directions (or, figuratively speaking, in antiphase):

- in the class of T1-oligomers, they decrease real values compared with model values of the hyperbolic progression 19569/n;
- in the class of G1-oligomers, they increase real values in comparison with model values of the hyperbolic progression 18901/n.

Secondly, under triplet-deviations, real total amounts of 3m-plets from the classes of T1-oligomers and G1-oligomers belong correspondingly to other hyperbolic progressions 11979/n and 28788/n. These hyperbolic progressions are indicated by the blue dot line and the green dot line in Fig. 8.2. Where did these numerators of model hyperbolas come from? Each of these numerators is associated with the total amount of triplets (n = 3) in an appropriate class of oligomers in this gene: the total amount of triplets starting with nucleotide T is equal to 3993, and the total amount of triplets starting with nucleotide G is equal to 9596. To calculate the first values of the model hyperbolas, each of these amounts of triplets must be tripled, giving the shown numerators 11979 and 28788.

Similar triplet-deviations exist in the OS-representations not only of the TNT gene but also of other long genes, prokaryotic genomes, and viruses in different degrees as the author has discovered in the analysis of a limited set of nucleotide series by the OS-method. In the genetic code system, triplets have an important meaning, which differs from other n-plets: they encode amino acids and punctuations of protein synthesis. One can believe that the phenomenon of the triplet-deviations is related to this special meaning of triplets. For this reason, the deeper analysis of triplet-deviations in different species can be useful to study the secrets of the genetic system and biological evolution.
Fig. 8.2 demonstrated the highly regular rhythmic triplet-deviations for \( n = 1, 2, 3, \ldots, 20 \), but similar rhythmic triplet-deviations exist in a much wider range of values \( n \). Fig. 8.3 shows in graphical forms percentage values of the highly regular rhythmic deviations of the real total amounts of \( n \)-plets, which start with the nucleotide \( T \) and with the nucleotide \( G \) in the \( TNT \) gene, from the appropriate model values \( 19569/n \) and \( 18901/n \). Two cases of the range of values \( n \) are represented there: \( n = 1, 2, 3, \ldots, 20 \), and \( n = 1, 2, 3, \ldots, 100 \).

Fig. 8.3. Percentage representations of highly regular rhythmic sequences of the triplet-deviations of the real amounts of \( n \)-plets, which belong to classes of \( T_1 \)- and \( G_1 \)-oligomers, from the appropriate model hyperbolic values \( 19569/n \) and \( 18901/n \) in the \( TNT \) gene. Here \( n = 1, 2, 3, \ldots, 20 \) (upper row) and \( n = 1, 2, 3, \ldots, 100 \) (bottom row) as shown at the abscissa axes. The ordinate axes show percentages of the deviations (the model values are taken as 100%).

The nucleobases \( T \) and \( G \) are keto-nucleobases. Figs. 8.2 and 8.3 draw attention to the phenomenon of long-range correlations in the \( TNT \) gene between sequences of the triplet-deviations in classes of \( T_1 \)- and \( G_1 \)-oligomers: the triplet-deviations in these sequences happen in opposite directions as above mentioned. Such binary oppositions, which meet in different long genes, prokaryotic genomes, and viruses regarding the classes of different \( N_1 \)-oligomers (here \( N \) refers to \( A, T, C, \) or \( G \)), should be specially studied in future since they bear important information and are associated with other binary-opposition features of molecular genetic systems.

The following Fig. 8.4 shows the OS-sequences of the total amounts of \( n \)-plets, which start with two other nucleotides \( A \) and \( C \) in the \( TNT \) gene. This gene contains 26373 nucleotides \( A \) and 17097 nucleotides \( C \); correspondingly \( S_A = 26373 \) and \( S_C = 17097 \) for the model hyperbolic progressions (2.2).
The class of $A_1$-oligomers

The class of $C_1$-oligomers

Fig. 8.4. Graphical representations of the results of the analysis - by the oligomer sums method – of the human TNT gene regarding the sequences of the total amounts of $n$-plets, which start with the nucleotide A (left) and the nucleotide C (right). Here $n = 1, 2, 3, \ldots, 20$ (at the abscissa axes). **Upper row:** the red lines refer to model hyperbolic progressions $S_A/n = 26373/n$ and $S_C/n = 17097/n$ correspondingly. The ordinate axes show the total amounts of appropriate $n$-plets. The class of $C_1$-oligomers has regular sequences of the significant triplet-deviations at $3m$-plets shown by the blue line. **Bottom row:** percentage representations of the sequences of deviations of the real total amounts of $n$-plets of these classes from the appropriate model hyperbolic values $26373/n$ and $17097/n$ (the ordinate axes show these percentages). The model values are taken as 100%.

One can see in Fig. 8.4 that the class of $C_1$-oligomers has regular sequences of the significant triplet-deviations at $3m$-plets shown by the blue line. The class of $A_1$-oligomers has not such regular sequences of significant deviations; besides, its deviations are essentially less than deviations in the class of $C_1$-oligomers. In the class of $A_1$-oligomers, the real and model values differ little from each other, and therefore, in Fig. 8.4, the red line of model values covers the line of real values.

Fig. 8.5 shows the numeric results of the analysis of the TNT gene by the oligomer sums method.

|   |  1  |  2  |  3  |  4  |  5  |  6  |  7  |  8  |  9  | 10  |
|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| A | Real | 26373 | 13334 | 8848 | 6656 | 5346 | 4463 | 3805 | 3315 | 2924 | 2724 |
|   | Model| 26373 | 13187 | 8791 | 6593 | 5275 | 4396 | 3768 | 3297 | 2930 | 2637 |
|   | $\Delta$% | 0 | -1.19 | -0.648 | -0.952 | -1.354 | -1.536 | -0.993 | -0.557 | 0.216 | -3.287 |
|   | Real | 19569 | 9677 | 3993 | 4857 | 3885 | 1964 | 2755 | 2426 | 1332 | 1943 |
|   | Model| 19569 | 9784.5 | 6523 | 4892 | 3914 | 3262 | 2796 | 2446 | 2174 | 1957 |
|   | $\Delta$% | 0 | 1.099 | 38.786 | 0.721 | 0.736 | 39.782 | 1.451 | 0.823 | 38.740 | 0.710 |
| C | Real | 17097 | 8522 | 4876 | 4199 | 3426 | 2431 | 2458 | 2101 | 1617 | 1707 |
|   | Model| 17097 | 8549 | 5699 | 4274 | 3419 | 2850 | 2442 | 2137 | 1900 | 1710 |
|   | $\Delta$% | 0 | 0.310 | 14.441 | 1.761 | -0.193 | 14.687 | -0.638 | 1.690 | 14.880 | 0.158 |
| G | Real | 18901 | 9437 | 9596 | 4773 | 3731 | 4798 | 2687 | 2400 | 3231 | 1820 |
|   | Model| 18901 | 9451 | 6300 | 4725 | 3780 | 3150 | 2700 | 2363 | 2100 | 1890 |
|   | $\Delta$% | 0 | 0.143 | -52.309 | -1.011 | 1.302 | -52.309 | 0.487 | -1.582 | -53.849 | 3.709 |
Fig. 8.5. Real and model values to the OS-representations of the classes of A₁-, T₁-, C₁-, and G₁-oligomers in the human TNT gene are shown for \( n = 1, 2, \ldots, 20 \). The real total amounts of \( n \)-plets, which start with a certain nucleotide (A, T, C, or G), are indicated jointly with their model values \( H_{A,1}(n) = \frac{26373}{n} \), \( H_{T,1}(n) = \frac{19569}{n} \), \( H_{C,1}(n) = \frac{17097}{n} \), and \( H_{G,1}(n) = \frac{18901}{n} \) (in red). The symbol \( \Delta \% \) refers to deviations of real values from model values in percent (the model values are taken as 100\%).

The coordinated deviations of all four OS-sequences from their model harmonic progressions can be conveniently represented by the general sequence of harmonic mean values, which are calculated for their four corresponding members at each fixed \( n \). Fig. 8.6 shows such a sequence for the human TNT gene. One can see the very regular rhythmic nature of this general sequence of harmonic mean values, expressively reflecting the phenomenon of agreed triplet-deviations under 3\( m \)-plets in this gene.

![Graph showing harmonic mean values](https://www.ncbi.nlm.nih.gov/nuccore/X83957)

**Fig. 8.6.** The sequence of harmonic mean values of agreed deviations of all four OS-sequences from their model harmonic progressions \( H_{A,1}(n) = \frac{26373}{n} \), \( H_{T,1}(n) = \frac{19569}{n} \), \( H_{C,1}(n) = \frac{17097}{n} \), and \( H_{G,1}(n) = \frac{18901}{n} \) in the human TNT gene. \( n = 1, 2, \ldots, 20 \) are plotted along the abscissa axes. The ordinate axes show harmonic mean values.

Let us show briefly, for comparison, also the OS-representations of two long human genes: NEB gene and SYNE1 gene.

Figs. 8.7 and 8.8 show graphs with the results of the NEB gene by the oligomer sums method. Initial data on this gene were taken from [https://www.ncbi.nlm.nih.gov/nuccore/X83957](https://www.ncbi.nlm.nih.gov/nuccore/X83957). This gene contains 7071 nucleotides A, 4478 nucleotides T, 4578 nucleotides C, and 4754 nucleotides G, that is \( S_A = \).
7071, $S_T = 4478$, $S_C = 4578$, and $S_G = 4754$ for the model hyperbolic progressions (2.2). It can be especially noted that, in this gene, the amounts of nucleotides A and T are significantly different (7071 and 4478), that is, the second Chargaff's rule on their approximate equality in long sequences is not fulfilled here since this nucleotide sequence is not enough long for the Chargaff’s rule.

Fig. 8.7. Graphical representations of the results of the analysis - by the oligomer sums method – of the human NEB gene: the sequences of the total amounts of $n$-plets, which start with the nucleotide A (left) and the nucleotide G (right) are shown. Here $n = 1, 2, 3,\ldots, 20$ (at the abscissa axes). **Upper row:** the red lines refer to model hyperbolic progressions $S_A/n = 7071/n$ (left) and $S_G/n = 4754/n$ (right). The ordinate axes show the total amounts of appropriate $n$-plets. The highly regular sequences of the significant triplet-deviations at 3$m$-plets shown by the green line (left) and the blue line (right). **Bottom row:** percentage representations of the sequences of deviations of the real total amounts of $n$-plets of these classes from the appropriate model hyperbolic values 7071/$n$ and 4754/$n$ (the ordinate axes show these percentages). The model values are taken as 100%.

Fig. 8.8. additionally draws attention to the phenomenon of long-range correlations in the NEB gene between sequences of the triplet-deviations in classes of $A_1$, and $G_1$-oligomers: the triplet-deviations in these sequences happen in opposite directions.
Fig. 8.8. Graphical representations of the results of the analysis - by the oligomer sums method – of the human NEB gene: the sequences of the total amounts of n-plets, which start with the nucleotide T (left) and the nucleotide C (right) are shown. Here \( n = 1, 2, 3, \ldots, 20 \) (at the abscissa axes). **Upper row:** the red lines refer to model hyperbolic progressions \( S_T/n = 4478/n \) (left) and \( S_C/n = 4578/n \) (right). The ordinate axes show the total amounts of appropriate n-plets. The weakly regular sequences of the significant triplet-deviations at 3m-plets shown by the green line (left) and the blue line (right). **Bottom row:** percentage representations of the sequences of deviations of the real total amounts of n-plets of these classes from the appropriate model hyperbolic values 4478/n and 4578/n (the ordinate axes show these percentages). The model values are taken as 100%.

By analogy with Fig. 8.6, Fig. 8.9 shows the sequence of harmonic mean values of agreed deviations of all four OS-sequences from their model harmonic progressions for the case of the human NEB gene. One can see the very regular rhythmic nature of this general sequence of harmonic mean values, reflecting the phenomenon of agreed triplet-deviations under 3m-plets in this gene.

Fig. 8.9. The sequence of harmonic mean values of agreed deviations of all four OS-sequences from their model harmonic progressions in the human NEB gene. \( n = 1, 2, \ldots, 20 \) are plotted along the abscissa axes. The ordinate axis shows harmonic mean values.

Figs. 8.10 and 8.11 show graphs with the results of the human SYNE1 gene by the oligomer sums method. Initial data on this gene were taken from [https://www.ncbi.nlm.nih.gov/nuccore/NM_182961](https://www.ncbi.nlm.nih.gov/nuccore/NM_182961). This gene contains 8697 nucleotides A, 6032 nucleotides T, 5940 nucleotides C, and 7039 nucleotides G, that is \( S_A = 8697, S_T = 6032, S_C = 5940, \) and \( S_G = 7039 \) for the model hyperbolic progressions (2.2). It can be especially noted that, in this gene, the amounts of nucleotides A and T are significantly different, as are the amounts of nucleotides C and G, and therefore the second Chargaff’s rule for long nucleotide sequences is not satisfied here since this nucleotide sequence is not enough long.
The class of $A_1$-oligomers

The class of $G_1$-oligomers

Fig. 8.10. Graphical representations of the results of the analysis - by the oligomer sums method – of the human SYNE1 gene: the sequences of the total amounts of $n$-plets, which start with the nucleotide A (left) and the nucleotide G (right) are shown. Here $n = 1, 2, 3,\ldots, 20$ (at the absciss axes). **Upper row:** the red lines refer to model hyperbolic progressions $S_A/n = 8697/n$ (left) and $S_G/n = 7039/n$ (right). The ordinate axes show the total amounts of appropriate $n$-plets. The highly regular sequences of the significant triplet-deviations at $3m$-plets shown by the blue line (left) and the green line (right). **Bottom row:** percentage representations of the sequences of deviations of the real total amounts of $n$-plets of these classes from the appropriate model hyperbolic values $7071/n$ and $4754/n$ (the ordinate axes show these percentages). The model values are taken as 100%.

Fig. 8.10 additionally draws attention to the phenomenon of long-range correlations in the SYNE1 gene between sequences of the triplet-deviations in classes of $A_1$-, and $G_1$-oligomers: the triplet-deviations in these sequences happen in opposite directions.
**Fig. 8.11.** Graphical representations of the results of the analysis - by the oligomer sums method – of the human *SYNE1* gene: the sequences of the total amounts of *n*-plets, which start with the nucleotide T (at left) and the nucleotide C (at right) are shown. Here *n* = 1, 2, 3, ..., 20 (at the absciss axes). **Upper row:** the red lines refer to model hyperbolic progressions \( S_T/n = 6032/n \) (left) and \( S_C/n = 5940/n \) (right). The ordinate axes show the total amounts of appropriate *n*-plets. The triplet-deviations in both of these classes are small in magnitude, and therefore, on these graphs, the model hyperbolic progressions (in red) practically hide under themselves the sequences of real total amounts of 3*m*-plets. **Bottom row:** percentage representations of the sequences of deviations of the real total amounts of *n*-plets of these classes from the appropriate model hyperbolic values 6032/*n* and 5940/*n* (the ordinate axes show these percentages). The model values are taken as 100%.

The author notes else that not all long genes have regular sequences of the pronounced triplet-deviations in their OS-representations. The comparison analysis of the OS-representations of different genes is a new research field.

One of the interesting topics for comparative analysis of genes by the oligomer sums method relates to the structure of histones, which is highly conservative in evolution. Histones are highly basic proteins found in eukaryotic cell nuclei that pack and order the DNA into structural units called nucleosomes. Histones are the chief protein components of chromatin, acting as spools around which DNA winds, and playing a role in gene regulation.

Figs. 8.12 and 8.13 show results of the analysis of one of the short genes of human histones by the OS-method (this gene was randomly selected from multiple histone genes for analysis): *H.sapiens* H1.1 gene for histone H1, 1034 bp DNA (GenBank: X57130.1, [https://www.ncbi.nlm.nih.gov/nuccore/X57130.1](https://www.ncbi.nlm.nih.gov/nuccore/X57130.1)). The results confirm the implementation of the hyperbolic (harmonic) rule № 1 for this gene.
Fig. 8.12. Graphical representations of the results of the analysis – by the oligomer sums method – of the human histone H1.1 gene regarding its classes of C₁- and G₁-oligomers. The sequences of its total amounts of n-plets, which start with the nucleotide C (left) and the nucleotide G (right), jointly with model hyperbolic progressions are shown. **Top graphic row:** the red lines refer to model hyperbolic progressions $S_C/n$ and $S_G/n$ correspondingly, where $S_C = 268$ and $S_G = 256$ are the quantities of nucleotides C and G in the gene; $n = 1, 2, 3, \ldots, 20$ as shown at the abscissa axes. The green line (left) and the blue line (right) with dots on them refer to the real OS-sequences of the total amounts of such n-plets. The ordinate axes indicate the total amounts of n-plets. **Middle graphic row:** the graph combining both graphs from the top row. **Bottom row:** graphs indicate the sequences of percent deviations of the real total amounts of n-plets, which start with appropriate nucleotides C and G in the gene, from the model hyperbolic values $S_C/n$ (left) and $S_G/n$ (right) under $n = 1, 2, 3, \ldots, 20$ (at the absciss axis). The ordinate axis indicates values of percent deviations. The model values are taken as 100%.

One can see in Fig. 8.12. the existence in this short histone gene some analog of those triplet-deviations related to $3m$-plets that were described above for long genes and shown in Figs. 8.1-8.10. In particular, the correlation exists in this short gene between two sequences of the triplet-deviations in the considered classes of C₁- and G₁-oligomers: the triplet-deviations in these sequences happen in opposite directions.

Fig. 8.13 shows the results of a similar analysis of the same histone gene regarding its classes of A₁- and T₁-oligomers.

![Graphical representations of the results of the analysis](https://example.com/graph.png)
Fig. 8.13. Graphical representations of the results of the analysis - by the oligomer sums method – of the human histone H1.1 gene regarding its classes of A₁- and T₁-oligomers. **Top graphic row:** the red lines with red dots refer to model hyperbolic progressions $S_A/n$ and $S_T/n$ correspondingly, where $S_A = 317$ and $S_T = 193$ are the quantities of nucleotides A and T; $n = 1, 2, 3, \ldots, 20$ as shown at the abscissa axes. The green line (left) and the blue line (right) with dots on them refer to the real OS-sequences of the total amounts of such $n$-plets. The ordinate axes indicate the total amounts of $n$-plets. **Bottom row:** graphs indicate the sequences of percent deviations of the real total amounts of $n$-plets, which start with appropriate nucleotides A and T in the gene, from the model hyperbolic values $S_A/n$ and $S_T/n$ under $n = 1, 2, 3, \ldots, 20$ (at the absciss axis). The ordinate axis indicates values of percent deviations. The model values are taken as 100%.

One can see from Fig. 8.13 that rhythmic deviations in the classes of A₁- and T₁-oligomers are less regular and stable than in the classes of C₁- and G₁-oligomers in Fig. 8.12. To clarify the general picture of such properties of histone genes, systematic studies of a wide set of histone genes in their OS-representations are required.

By analogy with Figs. 8.6 and 8.9, Fig. 8.14 shows the sequence of harmonic mean values of agreed deviations of all four OS-sequences from their model harmonic progressions for the case of the human histone H1.1 gene. One can see the regular rhythmic nature of this general sequence of harmonic mean values, reflecting the phenomenon of agreed triplet-deviations under $3m$-plets in this gene.

![Graph showing harmonic mean values](image)

**Fig. 8.14.** The sequence of harmonic mean values of agreed deviations of all four OS-sequences from their model harmonic progressions in the human histone H1.1 gene. $n = 1, 2, \ldots, 20$ are plotted along the abscissa axes. The ordinate axis shows harmonic mean values.

Certain triplet-deviations between real and model values under $3m$-plets are also found in the OS-representations of entire chromosomes of humans and other organisms, but in a much less pronounced form than shown in this Section for individual genes.

**9. The hyperbolic rules in bacterial genomes of different groups both from Bacteria and Archaea.**

Let us turn now to prokaryotic genomes. The Section represents results of the analysis of nucleotide sequences of all 19 bacterial genomes of different groups both from Bacteria and Archaea, which are listed in the article on the second Chargaff’s rule [Rapoport, Trifonov, 2012, p. 2]: “Nucleotide disparities for prokaryotic coding sequences were taken from bacterial genomes of different groups both from Bacteria and Archea. All together 19 genomes were used: Aquifex aeolicus, Acidobacteria...
bacterium, Bradyrhizobium japonicum, Bacillus subtilis, Chlamydia trachomatis, Chromobacterium violaceum, Dehalococcoides ethenogenes, Escherichia coli, Flavobacterium psychrophilum, Gloeobacter violaceus, Helicobacter pilory, Methanosarcina acetivorans, Nanoarchaeum equitans, Syntrophus aciditrophicus, Streptomyces coelicolor, Sulfolobus solfataricus, Treponema denticola, Thermotoga maritima and Thermus thermophiles”.

Fig. 9.1 shows the results of the analysis of these prokaryotic genomes by the oligomer sums method. These results demonstrate that the hyperbolic rule No. 1 is fulfilled for all the listed genomes of prokaryotes: the model hyperbolic progressions $H_{A,1}(n) = S_A/n$, $H_{T,1}(n) = S_T/n$, $H_{C,1}(n) = S_C/n$, and $H_{G,1}(n) = S_G/n$ from the expression (2.2) practically coincide with the OS-sequences of real total amounts of $n$-plets from the classes $A_1$-, $T_1$-, $C_1$-, and $G_1$-oligomers at $n = 1, 2, 3, \ldots, 20$. Because of this coincidence, the model hyperbolic progressions, which are represented by red lines in the graphs of Fig. 9.1, almost completely cover the sequences of real values (the blue lines in the lower graphs show in percent slight alternating deviations of real values from model values).
| Page | Data Values |
|------|-------------|
| 3    | $S_A = 1129118$, $S_T = 1129396$, $S_C = 882500$, $S_G = 884272$ |
| 4    | $S_A = 301793$, $S_T = 300618$, $S_C = 212019$, $S_G = 211409$ |
| 5    | $S_A = 21274$, $S_T = 23172$, $S_C = 38842$, $S_G = 44089$ |
| 6    | $S_A = 405227$, $S_T = 402383$, $S_C = 355663$, $S_G = 358014$ |
| Page | SA  | ST  | SC  | SG  |
|------|-----|-----|-----|-----|
| 8    | 1297551 | 1293044 | 1321325 | 1319228 |
| 9    | 945771  | 975318  | 468718  | 458213  |
| 10   | 887941  | 882586  | 1444547 | 1443945 |
| 11   | 498514  | 501121  | 323770  | 320426  |
\[ S_A = 867639 \]
\[ S_T = 881683 \]
\[ S_C = 490453 \]
\[ S_G = 487562 \]

\[ S_A = 570544 \]
\[ S_T = 572448 \]
\[ S_C = 346447 \]
\[ S_G = 353748 \]

\[ S_A = 501112 \]
\[ S_T = 498004 \]
\[ S_C = 424115 \]
\[ S_G = 436351 \]

\[ S_A = 327251 \]
\[ S_T = 330338 \]
\[ S_C = 734285 \]
\[ S_G = 729652 \]
Fig. 9.1. Graphical representations of the results of the analysis - by the oligomer sums method – of 19 bacterial genomes of Bacteria and Archaea mentioned in [Rapoport, Trifonov, 2012, p. 2]. For each of genomes two rows of resulting data are shown at n = 1, 2, …, 20 plotted along the abscissa axes: the top rows demonstrate that model hyperbolic progressions $S_A/n$, $S_T/n$, $S_C/n$, $S_G/n$ (red lines) almost completely cover the OS-sequences of real values (the ordinate axes show appropriate values); the bottom blue lines show in percent slight alternating deviations of real values from model values. The left column indicates numbers denoted the genomes as explained in the text.

The genomes are enumerated in Fig. 9.1 by numbers 1-19:

1) Aquifex aeolicus VF5, complete genome, 1551335 bp, accession AE000657, version AE000657.1, [https://www.ncbi.nlm.nih.gov/nuccore/AE000657.1?report=genbank](https://www.ncbi.nlm.nih.gov/nuccore/AE000657.1?report=genbank);

2) Acidobacteria bacterium KBS M015 of draft sequence, 4996384 bp, accession JHVA01000001, [https://www.ncbi.nlm.nih.gov/nuccore/JHVA01000001.1?report=genbank](https://www.ncbi.nlm.nih.gov/nuccore/JHVA01000001.1?report=genbank);

3) Bradyrhizobium japonicum strain E109, complete genome, 9224208 bp, accession CP010313, [https://www.ncbi.nlm.nih.gov/nuccore/CP010313.1?report=genbank](https://www.ncbi.nlm.nih.gov/nuccore/CP010313.1?report=genbank);

4) Bacillus subtilis strain UD1022, complete genome, 4025326 bp, accession CP011534, [https://www.ncbi.nlm.nih.gov/nuccore/CP011534.1?report=genbank](https://www.ncbi.nlm.nih.gov/nuccore/CP011534.1?report=genbank);

5) Chlamydia trachomatis strain QH111L, complete genome, 1025839 bp, accession CP018052, [https://www.ncbi.nlm.nih.gov/nuccore/CP018052.1?report=genbank](https://www.ncbi.nlm.nih.gov/nuccore/CP018052.1?report=genbank);

6) Chromobacterium violaceum strain LK30 1, whole genome shotgun sequence, 127377 bp, accession LDUX01000001, [https://www.ncbi.nlm.nih.gov/nuccore/LDUX01000001.1?report=genbank](https://www.ncbi.nlm.nih.gov/nuccore/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/nuccore/NZ_CP013074.1?report=genbank](https://www.ncbi.nlm.nih.gov/nuccore/NZ_CP013074.1?report=genbank);

8) Escherichia coli CFT073, complete genome, GenBank: AE014075.1, 5231428 bp, [https://www.ncbi.nlm.nih.gov/nuccore/AE014075.1?report=genbank](https://www.ncbi.nlm.nih.gov/nuccore/AE014075.1?report=genbank);

9) Flavobacterium psychrophilum JIP02/86, complete genome, 2860382 bp, accession NC_009613, [https://www.ncbi.nlm.nih.gov/nuccore/NC_009613.3](https://www.ncbi.nlm.nih.gov/nuccore/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/nuccore/BA000045.2?report=genbank](https://www.ncbi.nlm.nih.gov/nuccore/BA000045.2?report=genbank);

11) Helicobacter pilory, NCBI Reference Sequence: NC_00921.1, complete genome, 1643831 bp, accession NC_009211 NZ_AE001440-NZ_AE001571.
version NC_000921.1, https://www.ncbi.nlm.nih.gov/nuccore/NC_000921.1;
12) Methanosarcina acetivorans str. C2A, complete genome, 5751492 bp, accession AE010299-AE011189 version AE010299.1, https://www.ncbi.nlm.nih.gov/nuccore/AE010299;
13) Nanoarchaeum equitans Kin4-M, complete genome, 490885 bp, accession AE017199 AACL01000000 AACL01000001 version AE017199.1, https://www.ncbi.nlm.nih.gov/nuccore/AE017199.1?report=genbank;
14) Syntrophus aciditrophicus SB, complete genome, 3179300 bp, accession CP000252, https://www.ncbi.nlm.nih.gov/nuccore/CP000252.1?report=genbank;
15) Streptomyces coelicolor A3(2) complete genome, 8667507 bp, accession AL645882, https://www.ncbi.nlm.nih.gov/nuccore/AL645882.2?report=genbank;
16) Sulfolobus solfataricus strain SULA, complete genome, 2727373 bp, accession CP011057, https://www.ncbi.nlm.nih.gov/nuccore/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_AGDZ01000000 version NZ_KB442453.1, https://www.ncbi.nlm.nih.gov/nuccore/NZ_KB442453.1?report=genbank;
18) Thermotoga maritima strain Tma200, complete genome, 1859582 bp, accession CP010967, https://www.ncbi.nlm.nih.gov/nuccore/CP010967.1?report=genbank;
19) Thermus thermophilus DNA, complete genome, strain: TMY, 2121526 bp, accession AP017920, https://www.ncbi.nlm.nih.gov/nuccore/AP017920.1?report=genbank

One can see from Fig. 9.1 that in some prokaryotic genomes (for example in №№ 3, 7, 9, and 15) the alternating small deviations of real values from model values are systematic and related to 3m-plets; it seems to be analogous to the much stronger triplet-deviations described above for human genes in Figs. 8.1-8.7. Can a sign of the presence of such triplet-deviations in the genomes of some bacteria serve as a criterion for the selection of bacterial species for genetic engineering problems? It is one of many new questions arisen due to the discovery of the represented hyperbolic rules and the applications of the oligomer sums method.

Fig. 9.2 shows examples of sequences of the harmonic mean values for two of these bacterial genomes. One can see triplet-deviations in these sequences at dots corresponding to 3m-plets.
Fig. 9.2. The sequences of harmonic mean values of agreed deviations of all four OS-sequences from their model harmonic progressions in the genomes of Bradyrhizobium japonicum strain E109 (left) and Escherichia coli CFT073 (right). \( n = 1, 2, \ldots, 20 \) are plotted along the abscissa axes. The ordinate axes show harmonic mean values.

10. Analysis of genomes of microorganisms living in extreme environments

Of particular interest is the analysis of the genetic characteristics of microorganisms (extremophiles) living under extreme conditions of high and low temperatures, radiation, acidic and alkaline environments, drying, etc. Study of extremophiles is useful for many practical and theoretical problems. The https://en.wikipedia.org/wiki/Extremophile website contains a table of extremophiles. For the analysis of their genomes by the oligomer sums method, the author used 1-2 organisms from each category of the table. The initial data on the genomes were taken from the GenBank. Figs. 10.1-10.8 show the results of their analysis.

Fig. 10.1. The results of the analysis - by the oligomer sums method – the extremophile *Pyrolobus fumarii* 1A, complete genome, 1843267 bp (this extremophile lives in submarine hydrothermal vents), https://www.ncbi.nlm.nih.gov/nuccore/NC_015931.1. All abscissa axes show the values \( n = 1, 2, \ldots, 20 \). The top row demonstrates that model hyperbolic progressions \( S_A/n, S_T/n, S_C/n, S_G/n \) (red lines) almost completely cover the OS-sequences of real values (the ordinate axes show appropriate values). The bottom row show in percent slight alternating deviations of real values of the OS-sequences from model values.
**Fig. 10.2.** The results of the analysis - by the oligomer sums method – the extremophile *Pyrococcus furiosus DSM 3638*, complete genome, 1908256 (this extremophile lives in submarine hydrothermal vents). 
https://www.ncbi.nlm.nih.gov/nuccore/NC_003413.1 
The explanation of these graphs is identical to the explanation to Fig. 10.1.

**Fig. 10.3.** The results of the analysis - by the oligomer sums method – the extremophile *Synechococcus lividus PCC 6715* chromosome, complete genome, 2659739 bp (this extremophile lives in low temperature conditions), 
https://www.ncbi.nlm.nih.gov/nuccore/NZ_CP018092.1). The explanation of these graphs is identical to the explanation to Fig. 10.1.

**Fig. 10.4.** The results of the analysis - by the oligomer sums method – the extremophile *Psychrobacter alimentarius* strain *PAMC 27889* chromosome,
complete genome, 3332539 bp (this extremophile lives in soda lakes),
https://www.ncbi.nlm.nih.gov/nuccore/NZ_CP014945.1. The explanation of these
graphs is identical to the explanation to Fig. 10.1.

Fig. 10.5. The results of the analysis - by the oligomer sums method – the
extremophile *Clostridium paradoxum* JW-YL-7 = DSM 7308 strain JW-YL-7 ctg1,
whole genome shotgun sequence, 1855173 bp (this extremophile lives in volcanic
springs, acid mine drainage),
https://www.ncbi.nlm.nih.gov/nuccore/LSFY01000001.1 The explanation of
these graphs is identical to the explanation to Fig. 10.1.

Fig. 10.6. The results of the analysis - by the oligomer sums method – the
extremophile *Deinococcus radiodurans* R1 chromosome 1, complete sequence,
2648638 bp, (this extremophile lives in conditions of cosmic rays, X-rays,
radioactive decay), https://www.ncbi.nlm.nih.gov/nuccore/NC_001263.1. The
explanation of these graphs is identical to the explanation to Fig. 10.1.
The results of analysis - by the oligomer sums method – the extremophile *Halobacterium sp. NRC-1*, complete genome, 2014239 bp (this extremophile lives in conditions of high salt concentration), [https://www.ncbi.nlm.nih.gov/nuccore/NC_002607.1](https://www.ncbi.nlm.nih.gov/nuccore/NC_002607.1). The explanation of these graphs is identical to the explanation to Fig. 10.1.

The resulting data in Figs. 10.1- 10.8 shows the fulfillment of the hyperbolic rule № 1 of oligomeric sums for all studied and presented extremophiles. The extremal living conditions of these microorganisms do not affect the subordination of their genomes to the described hyperbolic (harmonic) rules of algebraic invariance, which are true for the genomes of other prokaryotes and eukaryotes.
11. Analysis of giant viruses by the oligomer sums method

This Section represents examples of studying genomes of different viruses by the oligomer sums method. The focus is on giant viruses (Figs. 11.1-11.4).

**Fig. 11.1.** The results of the analysis - by the oligomer sums method – the giant virus Pithovirus sibericum isolate P1084-T, complete genome, 610033 bp, NCBI Reference Sequence: NC_023423.1 [https://www.ncbi.nlm.nih.gov/nuccore/NC_023423.1](https://www.ncbi.nlm.nih.gov/nuccore/NC_023423.1). All abscissa axes show the values n = 1, 2, …, 20. The top row demonstrates that model hyperbolic progressions \( S_A/n, S_T/n, S_C/n, S_G/n \) (red lines) almost completely cover the OS-sequences of real values (the ordinate axes show appropriate values). The bottom row show in percent slight alternating deviations of real values of the OS-sequences from model values.

**Fig. 11.2.** The results of the analysis - by the oligomer sums method – the giant virus Acanthamoeba castellanii mamavirus strain Hal-V, complete genome, 1191693 bp, GenBank: JF801956.1, [https://www.ncbi.nlm.nih.gov/nuccore/JF801956.1](https://www.ncbi.nlm.nih.gov/nuccore/JF801956.1). The explanation of these graphs is identical to the explanation to Fig. 11.1.
Fig. 11.3. The results of the analysis - by the oligomer sums method – the giant virus Megaviruses chiliensis, complete genome, 1259197 bp, NCBI Reference Sequence: NC_016072.1, https://www.ncbi.nlm.nih.gov/nuccore/NC_016072.1. The explanation of these graphs is identical to the explanation to Fig. 11.1.

Fig. 11.4. The results of the analysis - by the oligomer sums method – the giant virus Cafeteria roenbergensis virus BV-PW1, complete genome, 617453 bp, NCBI Reference Sequence: NC_014637.1, https://www.ncbi.nlm.nih.gov/nuccore/NC_014637.1. The explanation of these graphs is identical to the explanation to Fig. 11.1.

The results, presented in this Section, show the fulfillment of the hyperbolic (harmonic) rule No. 1 for the viruses considered and provide material for comparative analysis of different OS-sequences.

11. Analysis of the COVID-19 virus by the oligomer sums method

Let us turn now to the analysis - by the oligomeric sums method - of the COVID-19 virus, which led to a pandemic. The initial data on its nucleotide sequence was taken by the author from the site https://www.ncbi.nlm.nih.gov/nuccore/MN908947.3, where the following is written about it: severe acute respiratory syndrome coronavirus isolate Wuhan-Hu-1, complete genome, GenBank: MN908947.3, LOCUS MN908947, 29903 bp ss-RNA linear VRL 18-MAR-2020.
Figs. 12.1-12.5 show some results of such an analysis of the virus.

![Graph 1](image1)

**S_A = 8954**

**Fig. 12.1.** The graphs for the case of the OS-sequences of *n*-plets from the class A1-oligomers of the coronavirus 2 isolate Wuhan-Hu-1, complete genome, GenBank: MN908947.3, LOCUS MN908947, 29903 bp. In these graphs, the abscissa axis represents the values *n* = 1, 2, 3, ..., 20 (in top row) and *n* = 1, 2, 3, ..., 100 (in bottom row). **Top left:** the ordinate axis represents the set of phenomenological total amounts $\Sigma_{A,n,1}$ of *n*-plets beginning with the nucleotide A. **Top right, and bottom:** deviations of real OS-sequences $\Sigma_{A,n,1}$ from model hyperbolic progressions $S_A/n = 8954/n$ in percentages. 

![Graph 2](image2)

**S_T = 9594**
**Fig. 12.2.** The graphs for the case of the OS-sequences of \(n\)-plets from the class \(T_1\)-oligomers of the coronavirus 2 isolate Wuhan-Hu-1, complete genome, GenBank: MN908947.3, LOCUS MN908947, 29903 bp. In these graphs, the abscissa axis represents the values \(n = 1, 2, 3, \ldots, 20\) (in top row) and \(n = 1, 2, 3, \ldots, 100\) (in bottom row). **Top left:** the ordinate axis represents the set of phenomenological total amounts \(\Sigma_{T,n,1}\) of \(n\)-plets beginning with the nucleotide \(T\). **Top right, and bottom:** deviations of the real OS-sequence \(\Sigma_{T,n,1}\) from the model hyperbolic progression \(S_T/n = 9594/n\) in percentages.

\[ S_C = 5492 \]

\[ S_C = 5492 \]

\[ S_C = 5492 \]

**Fig. 12.3.** The graphs for the case of the OS-sequences of \(n\)-plets from the class \(C_1\)-oligomers of the coronavirus 2 isolate Wuhan-Hu-1, complete genome, GenBank: MN908947.3, LOCUS MN908947, 29903 bp. In these graphs, the abscissa axis represents the values \(n = 1, 2, 3, \ldots, 20\) (in top row) and \(n = 1, 2, 3, \ldots, 100\) (in bottom row). **Top left:** the ordinate axis represents the set of phenomenological total amounts \(\Sigma_{C,n,1}\) of \(n\)-plets beginning with the nucleotide \(C\). **Top right, and bottom:** deviations of the real OS-sequences \(\Sigma_{C,n,1}\) from the model hyperbolic progression \(S_C/n = 5492/n\) in percentages.

\[ S_G = 5863 \]

\[ S_G = 5863 \]
Fig. 12.4. The graphs for the case of the OS-sequences of \( n \)-plets from the class \( G_1 \)-oligomers of the coronavirus 2 isolate Wuhan-Hu-1, complete genome, GenBank: MN908947.3, LOCUS MN908947, 29903 bp. In these graphs, the abscissa axis represents the values \( n = 1, 2, 3, \ldots, 20 \) (in top row) and \( n = 1, 2, 3, \ldots, 100 \) (in bottom row). **Top left:** the ordinate axis represents the set of phenomenological total amounts \( \Sigma_{G,n,1} \) of \( n \)-plets beginning with the nucleotide \( G \). **Top right, and bottom:** deviations of the real OS-sequence \( \Sigma_{G,n,1} \) from the model hyperbolic progression \( S_{G/n} = 5863/n \) in percentages.

In particular, Figs. 12.1-12.4 show that this virus in its OS-representations has under \( n = 3, 6, 9, \ldots, 3m \) such deviations of real values from model values, which resemble the triplet-deviations in human genes, which were described above in Figs. 8.1-8.10. Perhaps the harmfulness of this virus to humans is related to this similarity. It should also be noted that - in the classes of pyrimidines \( C_1 \)- and \( T_1 \)-oligomers (Figs. 12.2 and 12.3) - these deviations occur in opposite directions in a coordinated manner, which indicates a particular consistency in the structure of the nucleotide sequence of this virus concerning pyrimidines classes.

By analogy with Figs. 8.6, 8.9, and 8.14, Fig. 12.5 shows the sequence of harmonic mean values of agreed deviations of all four OS-sequences from their model harmonic progressions for the case of the coronavirus 2 isolate Wuhan-Hu-1. Two cases are shown: for \( n = 1, 2, \ldots, 20 \) and \( n = 1, 2, \ldots, 100 \). One can see the regular rhythmic nature of this general sequence of harmonic mean values, reflecting the phenomenon of agreed triplet-deviations under \( 3m \)-plets in this coronavirus.

Fig. 12.5. The sequence of harmonic mean values of agreed deviations of all four OS-sequences from their model harmonic progressions in the coronavirus 2 isolate Wuhan-Hu-1. The ordinate axes show harmonic mean values. The left and right graphs show the cases of \( n = 1, 2, \ldots, 20 \) and \( n = 1, 2, \ldots, 100 \), which are plotted along the abscissa axes.
13. DNA epi-chains and the hyperbolic rules for oligomer sums

This Section presents some results of the study of special subsequences of long nucleotide sequences in single-stranded DNA by the oligomer sums method. These subsequences are termed «DNA epi-chains» [Petoukhov, 2019a]. The author's initial results testify that the above described hyperbolic rules of oligomer sums for genomes are also fulfilled for these epi-chains; it gives new materials to a known theme of fractal-like structures in genetics.

By definition, in a nucleotide sequence $N_1$ of any DNA strand with sequentially numbered nucleotides 1, 2, 3, 4, ... (Fig. 13.1a), epi-chains of different orders $n$ are those subsequences that contain only those nucleotides, whose numeration differ from each other by natural number $n = 1, 2, 3, ...$. For example, in any single-stranded DNA, epi-chains of the second order are two nucleotide subsequences $N_{2/1}$ and $N_{2/2}$ in which their nucleotide sequence numbers differ by $n = 2$: the epi-chain $N_{2/1}$ contains nucleotides with odd numerations 1, 3, 5, ... (Fig. 13.1b), and the epi-chain $N_{2/2}$ contains nucleotides with even numerations 2, 4, 6, ... (Fig. 13.1c). By analogy, epi-chains of the third order are those three nucleotide subsequences $N_{3/1}$, $N_{3/2}$, and $N_{3/3}$, each of which has sequence numbers that differ by $n = 3$: these epi-chains contain nucleotides with numerations 1, 4, 7, ... or 2, 5, 8, ... or 3, 6, 9, ..., respectively (Figs. 13.1d-f). The epi-chain of the first order $N_1$ coincides with the nucleotide sequence of the DNA strand (Fig. 13.1a).
**Fig. 13.1.** A schematic representation of a single-stranded DNA and its initial epi-chains of nucleotides, denoted by black circles. **a**, a sequence $N_1$ of numera
ted nucleotides of the DNA strand. **b**, an epi-chain of the second order $N_{2/1}$ beginning with nucleotide number 1. **c**, an epi-chain of the second order $N_{2/2}$ beginning with nucleotide number 2. **d**, an epi-chain of the third order $N_{3/1}$ beginning with nucleotide number 1. **e**, an epi-chain of the third order $N_{3/2}$ beginning with nucleotide number 2. **f**, an epi-chain of the third order $N_{3/3}$ beginning with nucleotide number 3.

The term "epi-chain" was coined from the Ancient Greek prefix epi-, implying features that are "on top of" DNA strands. In any DNA strand, each nucleotide belongs to many epi-chains having different orders $k$. The symbol “N” in the designation of DNA epi-chains corresponds to the first letter in the word “nucleotides”. In the designation “$N_{k/m}$” of single-stranded DNA epi-chains, the numerator "$k$" in the index indicates the order of the epi-chain, and the denominator "$m$" indicates the numeration of the initial nucleotide of this epi-chain along the DNA strand (Fig. 13.1a). For example, the symbol $N_{3/2}$ refers to the epi-chain of the third order with the initial nucleotide having the number 2 in the DNA strand: 2-5-8-… (Fig. 13.1e).

Each DNA epi-chain of $k$-th order (if $k = 2, 3, 4, ...$) contains $k$ times fewer nucleotides than the DNA strand and has its own arrangements of nucleobases A, T, C, and G. Each DNA epi-chain of the order $k$ (if $k = 2, 3, 4, ...$) contains $k$ times fewer nucleotides than the DNA strand and has its own arrangements of nucleobases A, T, C and G. But unexpectedly, despite on these differences, OS-sequences of the total amounts of those $n$-plets, which start with a nucleotide A, or T, or C, or G, are modeled by very similar hyperbolic progressions as in the complete DNA strand and as in its epi-chains (at this stage of the research, the author studied OS-representations of epi-chains only in cases of epi-chains with relatively small orders $k$).

Figs. 13.2-13.6 explains these results in graphical forms by examples of the OS-representations of epi-chains $N_{2/1}$, $N_{3/1}$, $N_{4/1}$, $N_{10/1}$, and $N_{50/1}$ in the human chromosome №1 (the OS-representation of this complete chromosome was presented above in Figs. 2.1-2.3).
Fig. 13.2. The results of the analysis - by the oligomer sums method – the nucleotide sequence of the epi-chain of the second order $N_{2/1}$ (Fig. 13.1b), which consists of nucleotides with serial numerations 1-3-5-7-9-... in the DNA sequence of the human chromosome № 1. All abscissa axes show the values $n = 1, 2, \ldots, 20$. The top row demonstrates that the model hyperbolic progressions $S_A/n$, $S_T/n$, $S_C/n$, $S_G/n$ (red lines) almost completely cover the OS-sequences of real total amounts of those $n$-plets, which start with a nucleotide A, or T, or C, or G in this epi-chain correspondingly (the ordinate axes show appropriate amounts). The bottom row show in percent slight alternating deviations of real values of the OS-sequences from model values.

Fig. 13.3. The results of the analysis - by the oligomer sums method – the nucleotide sequence of the epi-chain of the third order $N_{3/1}$ (Fig. 13.1d), which consists of nucleotides with serial numerations 1-4-7-10-13-... in the DNA sequence of the human chromosome № 1. The top row demonstrates that the model hyperbolic progressions $S_A/n$, $S_T/n$, $S_C/n$, $S_G/n$ (red lines) almost completely cover the OS-sequences of real total amounts of those $n$-plets, which start with a nucleotide A, or T, or C, or G in this epi-chain correspondingly. The bottom row show in percent slight alternating deviations of real values of the OS-sequences from model values. All denotations are the same as in Fig. 13.2.
Fig. 13.4. The results of the analysis - by the oligomer sums method – the nucleotide sequence of the epi-chain of the $4^{th}$ order $N_4/1$, which consists of nucleotides with serial numerations 1-5-9-13-… in the DNA sequence of the human chromosome № 1. The top row demonstrates that the model hyperbolic progressions $S_A/n$, $S_T/n$, $S_C/n$, $S_G/n$ (red lines) almost completely cover the OS-sequences of real total amounts of those $n$-plets, which start with a nucleotide A, or T, or C, or G in this epi-chain correspondingly. The bottom row show in percent slight alternating deviations of real values of the OS-sequences from model values. All denotations are the same as in Fig. 13.2.

Fig. 13.5. The results of the analysis - by the oligomer sums method – the nucleotide sequence of the epi-chain of the $10^{th}$ order $N_{10}/1$, which consists of nucleotides with serial numerations 1-11-21-31-41-… in the DNA sequence of the human chromosome № 1. The top row demonstrates that the model hyperbolic progressions $S_A/n$, $S_T/n$, $S_C/n$, $S_G/n$ (red lines) almost completely cover the OS-sequences of real total amounts of those $n$-plets, which start with a nucleotide A, or T, or C, or G in this epi-chain correspondingly. The bottom row show in percent slight alternating deviations of real values of the OS-sequences from model values. All denotations are the same as in Fig. 13.2.
Fig. 13.6. The results of the analysis - by the oligomer sums method – the nucleotide sequence of the epi-chain of the 50th order N_{50/1}, which consists of nucleotides with serial numerations 1-51-101-151-201— in the DNA sequence of the human chromosome № 1. The top row demonstrates that the model hyperbolic progressions $S_A/n$, $S_T/n$, $S_C/n$, $S_G/n$ (red lines) almost completely cover the OS-sequences of real total amounts of those n-plots, which start with a nucleotide A, or T, or C, or G in this epi-chain correspondingly. The bottom row show in percent slight alternating deviations of real values of the OS-sequences from model values. All denotations are the same as in Fig. 13.2.

Figs. 13.2-13.6 show that in these epi-chains, which are sparse subsequences of the complete DNA sequence, the same hyperbolic rule No. 1 is fulfilled, which was formulated above for complete DNA sequences in eukaryotic and prokaryotic genomes. The rule is fulfilled in these epi-chains with the same high accuracy as in the complete DNA of the sequence.

Similar results were obtained by the author in study of epi-chains in the single-stranded DNA of other analyzed genomes (see some corresponding data in [Petoukhov, 2019a]). These results allow formulating the fourth hyperbolic (or harmonic) rule of eukaryotic and prokaryotic genomes, which is considered by the author as a candidate for the role of a universal genetic rule (it is necessary to further investigate the widest variety of genomes to verify a degree of its universality).

**The fourth hyperbolic rule** (about interrelations of oligomers in epi-chains of long DNA sequences):

- In any nuclear chromosome of eukaryotic genomes and in prokaryotic genomes, the hyperbolic rules №№ 1 and 2 are fulfilled not only for the complete nucleotide sequences but also for their epi-chains of the order $k$ (where $k = 2, 3, 4, \ldots$ is not too large compared to the full length of the nucleotide sequence).

Appendix I shows the numeric data represented in the graphs in Figs. 13.2.-13.6.
14. The quantum-information model of the oligomer cooperative organization in genomes and its confirmed predictions

The Section is devoted to the connections of the described phenomenological hyperbolic (harmonic) rules in genomes with the concepts and mathematical formalisms of quantum informatics.

One of the creators of quantum mechanics P. Jordan in his work on quantum biology claimed that life's missing laws were the rules of chance and probability of the quantum world [Jordan, 1932; McFadden, Al-Khalili, 2018]. From the standpoint of Jordan’s statement, the study of probabilities or frequencies of n-plets (monoplets, doublets, triplets, etc.) in long DNA sequences is important for discovering hidden biological laws and for developing quantum biology. The phenomenological hyperbolic rules about the total amounts of certain oligomers in the genomes described above allow us to study their connection with the probability rules of these groups of oligomers in the genomes. Let us explain this.

Till now we considered the total amounts $\Sigma_{N,n,1}$ of certain n-plets, which start with the first nucleotide N (A, T, C, or G), and we discovered that, in different genomes, these amounts correspond to hyperbolic OS-sequences $S_N/n$ with a high accuracy, where $S_N$ refers to the total number of the nucleotide N. The whole sequence of nucleotides in a long single-stranded DNA can be considered as a sequence of oligomers of a certain length $n$, whose amount is equal to $S/n$. Each such oligomer starts with one of four nucleotides A, T, C, or G. Therefore the total amount $S/n$ of consecutive oligomers of length $n$ in the analyzed DNA sequence is the sum of all oligomers of length $n$ starting with A, or T, or C, or G:

$$S/n = \Sigma_{A,n,1} + \Sigma_{T,n,1} + \Sigma_{C,n,1} + \Sigma_{G,n,1}$$ \hspace{1cm} (14.1)

The collective probability (percentage, or frequency) $P_s(N_1)$ of all $\Sigma_{N,n,1}$ n-plets starting with the nucleotide N, relative to the amount $S/n$ (14.1), is determined by the expression (14.2):

$$P_s(N_1) = \Sigma_{N,n,1}/(S/n) \approx (S_N/n)/(S/n) = S_N/S = P(N)$$ \hspace{1cm} (14.2)

The expression (14.2) shows that the collective probability $P_s(N_1)$ is independent of $n$ and is approximately equal to the probability (frequency) $P(N) = S_N/S$ of the nucleotide N in the genomic sequence having S nucleotides.

For example, the human chromosome №1, which was considered above in Section 1 (Figs. 2.1-2.3), has the total amount of nucleotides $S = S_A + S_T + S_C + S_G = 67070277 + 67244164 + 48055043 + 48111528 = 230481012$. The probability $P(A)$ of the nucleotide A is equal to $S_A/S = 67070277 / 230481012 \approx 0.2910$. From the data in Fig. 2.3, one can verify that, in this chromosome, the collective probabilities $P_s(A_1)$ of total amounts of n-plets ($n = 2, 3, \ldots, 20$) starting with the nucleotide A are also equal to this value $P(A) = 0.2910$ with a high level of accuracy independently of $n$. A similar situation holds with respect to the nucleotides T, C, and G.

It is also useful to note the opposite: if, for a genome, the phenomenological probabilities of n-plets $P_n(N_1)$ (where $n = 1, 2, 3, \ldots$) are initially known, and their compliance with the rule - of type $P(N) \approx P_s(N_1)$ - of approximate equality of collective probability of n-plets is also known, then connection (14.2) allows us to construct a hyperbolic OS-sequence of the sums $\Sigma_{N,n,1}$ of n-plets (14.3):
This is noted here because the author previously discovered and published [Petoukhov, 2018b] the rules of the approximate equality of the collective probabilities of \( n \)-plets for \( n = 1, 2, 3, ... \). Given the expressions (14.2) and (14.3), the hyperbolic rules of the OS-sequences and these rules for the approximate equality of the collective probabilities of \( n \)-plets are equivalent. Both of them reflect in different languages the oligomeric cooperative organization of genomes. This is useful to note because the author has published an effective mathematical model for the rules of collective probability, which is obviously applicable also to the above formulated hyperbolic rule № 1 [Petoukhov, 2018b; Petoukhov, Petukhova, Svirin, 2019].

One should emphasize the following important aspect of the OS-representations of genomic sequences. Each nucleotide of a DNA sequence is a participant of those sets of its different \( n \)-plets (doublets, triplets, etc.), whose total amounts are members of OS-sequences of this DNA; in other words, each DNA nucleotide makes its small contribution immediately to many members of the OS-sequences. Figuratively speaking, each DNA nucleotide is "smeread" (or distributed) over many members of the DNA OS-sequence (this "smearing" over many members of the OS-sequence is also true for each DNA doublet, triplet, etc.). Correspondingly, OS-sequences reflect a sort of an interrelation over all \( n \)-plets in DNA sequences. Or, in other words, the oligomer sums method represents any long nucleotide sequence as a multi-partite (or many-body) system having a cooperative state regarding many its interrelated oligomers of different lengths \( n = 1, 2, 3, ... \).

This has some analogies with the well-known problem of multi-partite entanglement in quantum informatics described, for example, in [Walter, Gross, Eisert, 2017; Horodecki, Horodecki, et al., 2009; Gühne, Tóth, 2009; Amico, Fazio, et al., 2008].

Quantum entanglement is the physical phenomenon that occurs when a pair or group of particles is generated, interact, or share spatial proximity in a way such that the quantum state of each particle of the pair or group cannot be described independently of the state of the others. In quantum informatics, entangled states play very important roles. The study and use of entangled states are one of the main problems of quantum computing: "...entanglement is a key element in effects such as quantum teleportation, fast quantum algorithms, and quantum error-correction. It is, in short, a resource of great utility in quantum computation and quantum information. ...entangled states play a crucial role in quantum computation and quantum information" [Nielsen, Chuang, 2010, p. XXIII and p. 96].

Quantum systems with many degrees of freedom are ubiquitous in nature, particularly in the context of condensed matter theory. "It is hence not surprising that important classes of states, such as ground states of local Hamiltonians, are multi-partite entangled states. ...Recent years have seen an enormous increase in interest at the intersection of quantum information and condensed matter theory that stems from the insight that notions of entanglement are crucial in the understanding of quantum phases of matter .... Another family of quantum many-body states that can be efficiently described is the classes of bosonic and fermionic Gaussian states. They both arise naturally in the context of quantum many-body models in condensed matter physics, but their bosonic variant is also highly useful in quantum optics when it comes to describing systems constituted of several quantum modes of light... . Relatively, multi-partite entangled states serve as resources to a number of important
protocols in quantum information theory in which more than two parties come together. A prominent example of such a multi-party quantum protocol is quantum secret sharing, in which a message is distributed to several parties in such a way that no subset is able to read the message, but the entire collection of parties is. 

Multi-partite entanglement does not only facilitate processing or transmission of information but also allow for applications in metrology” [Walter, Gross, Eisert, 2017, pp. 15, 18, 20, 23]. The entanglement refers to the nonlocal properties of quantum states that cannot be explained classically.

Distinguish entanglement of distinguishable and indistinguishable (identical) particles. The state of the system K of distinguishable particles in a pure state is determined by the state vector \(|\psi\rangle\) in the Hilbert space \(H\), which is the tensor product of the subspaces corresponding to each particle:

\[
H = H_1 \otimes H_2 \otimes \ldots \otimes H_K \tag{14.3}
\]

If the particles are not entangled, then the state of the system is defined as the tensor product of the state vectors \(|\psi^{(i)}\rangle\) of the subsystems:

\[
|\psi\rangle = |\psi^{(1)}\rangle \otimes |\psi^{(2)}\rangle \otimes \ldots \otimes |\psi^{(K)}\rangle \tag{14.4}
\]

If the vector cannot be expressed in this form (14.4), then they say that the particles are quantum entangled.

The tensor product gives a way of putting separate vector spaces together to form larger vector spaces and it is one of the basis instruments in quantum informatics. The following quotation speaks about the meaning of the tensor product: “This construction is crucial to understanding the quantum mechanics of multiparticle systems ” [Nielsen, Chuang, 2010, p. 71]. But above Section 3 described that the DNA alphabets of 4 nucleotides, 16 doublets, 64 triplets,…, \(4^n\) n-plets, which have binary-oppositional systems of molecular traits, are interrelated by the tensor product of matrices representing them: these genetic matrices of DNA alphabets are members of a single tensor family \([G, T; C, A]^{(n)}\) (Fig. 3.1). This fact is one of the arguments in favor of the adequacy of the quantum-information approach to the study of genetic informatics and living bodies as informational entities.

One can suppose that in eukaryotic and prokaryotic genomes we have some special case of multi-partite entangled states, but not in groups of many particles, but in genomic systems of many oligomers. This can be termed as “the genomic entanglement” or as “the genomic tetra-entanglement” since genomic sequences contain 4 kinds of nucleotides A, T, C, and G.

Let us turn to the above-mentioned author's model of properties of genomic sequences expressed by the expressions (14.2) and (14.3) [Petoukhov, 2018b; Petoukhov, Petukhova, Svirin, 2019]. This model is based on the tensor products and some other formalisms of quantum informatics and concerns, first of all, the hyperbolic rule №1 of the oligomer cooperative organization of genomes. The model introduced the notion “genetic qubits” based on different pairs of binary-oppositional molecular traits of adenine A, guanine G, cytosine C, and thymine T. Appropriate \(2^n\)-qubit systems in separable pure states were constructed, where nucleotides A, T, C, and G (and also DNA doublets and other n-plets) were represented by appropriate computational basis states in Hilbert spaces of corresponding dimensionalities. For
example, cytosine C was represented as the computational basis state |00> of the 2-qubit system in the 4-dimensional Hilbert space, thymine T as the computational basis state |01>, guanine G as the computational basis state |10>, and adenine A as the computational basis state |11> of the same 2-qubit system. Correspondingly, 16 doublets were represented as 16 computational basis states of the 4-qubit system in the 16-dimensional Hilbert space: for example, the doublet CC was represented as the computational basis state |0000>, the doublet CT as |0001>, …, etc. This model can be used for a deeper understanding of the genomic entanglement.

An effective model should not only explain known phenomenological data but also predict unknown data to search them in natural systems. Let us show now that the proposed quantum-informational model has predictive power, allowing us to open previously unknown properties of genomic DNA sequences. Really, the noted model allowed a prediction not only the hyperbolic rule №1 described above but also many other non-trivial interrelations in genomic structures. In a limited volume of this article, the author can show only a few following brief examples.

**About additional confirmations of the model predictions.** For example, the model predicts the following. Till now we considered OS-sequences, whose members are total amounts of n-plets, which start with a certain «attributive» nucleotide, for example, with the nucleotide A. In this case, we calculate the total amounts of oligomers in the following sets: 4 doublets AT, AC, AG, AA; 16 triplets ATT, ATC, ATG, ACC, …; and so on. But what results arise if one calculates, in the same genome, the total amounts in quite other sets of n-plets having the same attributive nucleotide A at their second positions, that is the following sets: 4 doublets TA, CA, GA, AA; 16 triplets TAT, TAC, TAG, CAC, …; and so on for \( n = 2, 3, 4, \ldots \)? And what results arise if one calculates, in the same genome, total amounts in the sets of n-plets, which have the same nucleotide at their third positions, that is the following sets: 16 triplets TTA, TCA, TGA, CCA, …; 64 tetraplets TTTA, TCTA, TGCA, …; and so on for \( n = 3, 4, 5, \ldots \)? The quantum-information model predicts that in all such cases the resulting OS-sequences will be practically identical to the hyperbolic-like OS-sequence of the total amounts of n-plets with the same attributive nucleotide at their first position. These model predictions also apply to cases of sets of n-plets, which have the same attributive nucleotide at their 4th, 5th, 6th, …, kth positions for \( n = k, k+1, k+2, \ldots \) (here \( k \) is not too large compared to the full length of the genomic sequence).

These model predictions are confirmed by direct calculations of total amounts of corresponding sets of n-plets in different genomes. Figs. 14.1 and 14.2 show examples of such confirmations by the comparisons of different OS-sequences calculated for the human chromosome №1 in three cases of locations of attributive nucleotides in its n-plets: 1) at the first position in n-plets (data on the appropriate OS-sequences are taken from Fig. 2.3); 2) at the second position; 3) at the third position.

One can see from the shown results that the differences \( \Delta\% \) of the corresponding members of these three OS-sequences from each other are less than 0.1%, that is these OS-sequences are practically identical. These differences were calculated for each \( n \) by the formulas \( \Delta\% = 100(1 - \text{Pos1/Pos2})\% \) and \( \Delta\% = 100(1 - \text{Pos1/Pos3})\% \) where Pos1, Pos2, and Pos3 refer to values indicated in the rows Pos. 1, Pos. 2, and Pos. 3. Here the results are presented only for \( n = 2, 3, 4, \ldots, 10 \) but similar situations of practical coincidences of the corresponding members of the considered OS-sequences are also true for larger \( n \).
**Fig. 14.1.** The comparison of the OS-sequences of the total amounts of $n$-plets, which have the nucleotide N (A, T, C, or G) at their first position (the row “Pos. 1”) and at their second position (the row “Pos. 2”) in the human chromosome №1. Δ% shows the percentage of differences between the corresponding total amounts of $n$-plets from each other. The comparison begins with doublets, since there is no second position in monoplets.

| $n$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|-----|---|---|---|---|---|---|---|---|---|----|
| A   | 67070277 | 33537501 | 22360413 | 16768845 | 13413532 | 11719286 | 11584038 | 8383461 | 7453552 | 6706672 |
| Pos. 1 | 33532177 | 22353979 | 16767465 | 13413514 | 11714459 | 11578138 | 8383936 | 7452356 | 6704047 |
| Pos. 2 | 0.104 | 0.092 | 0.088 | 0.080 | 0.072 | 0.062 | 0.096 | 0.016 | 0.039 |
| T   | 67244164 | 33620498 | 22412993 | 16808862 | 13345360 | 11202724 | 9566748 | 8405040 | 7470145 | 6724359 |
| Pos. 1 | 33623666 | 22411661 | 16811071 | 13345910 | 11206100 | 9560249 | 8405351 | 7472348 | 6724456 |
| Pos. 2 | 0.09 | -0.008 | 0.013 | 0.004 | -0.010 | 0.036 | 0.004 | 0.029 | 0.001 |
| C   | 48055043 | 24024903 | 16012711 | 12013624 | 9612227 | 8005708 | 6865944 | 6008215 | 5336968 | 4803919 |
| Pos. 1 | 24030140 | 16014644 | 12015843 | 9615911 | 8012553 | 6865662 | 6005968 | 5338638 | 4808410 |
| Pos. 2 | 0.032 | 0.055 | 0.018 | 0.038 | 0.085 | -0.004 | -0.037 | 0.031 | 0.093 |
| G   | 48111528 | 24057606 | 16040889 | 12028924 | 9625086 | 8021235 | 6891932 | 6013412 | 5348337 | 4813156 |
| Pos. 1 | 24053922 | 16040412 | 12025875 | 9620866 | 8020389 | 6871831 | 6014853 | 5345656 | 4811187 |
| Pos. 2 | -0.015 | -0.003 | -0.025 | -0.044 | -0.011 | 0.039 | 0.024 | -0.050 | -0.041 |

**Fig. 14.2.** The comparison of the OS-sequences of the total amounts of $n$-plets, which have the nucleotide N (A, T, C, or G) at their first position (the row “Pos. 1”) and at their third position (the row “Pos. 3”) in the human chromosome №1. Δ% shows the percentage of differences of the corresponding total amounts of $n$-plets from each other. The comparison begins with triplets since there is no third position in monoplets and doublets.
These predictions about the oligomer cooperative organization and their confirmations in eukaryotic and prokaryotic genomes give a significant extension to the hyperbolic rule №1 regarding the hyperbolic-like OS-sequences of the total amounts of $n$-plets, which have the same attributive nucleotide at their $k$th position (not only in their first position). These results and the extended rules additionally open up the deep connections of genomic sequences with the harmonic progression (2.4) and discover new aspects of the algebraic harmony of living bodies.

Another large bunch of predictions about genomic sequences is given by the quantum-information model for quantitative interrelations of different $n$-plets, which start from the same doublet, or from the same triplet, etc. The model predicts, in particular, that the amount $S_2$ of any of 16 doublets NN is algebra-harmonically interrelated with the total amounts $S_3$, $S_4$, $S_5$, ... of oligomers in the following sets: 4 triplets, which start with this attributive doublet NN; 16 tetraplets, which start with this attributive doublet NN; 64 pentaplets, which start with this attributive doublet NN; and so on. This interrelation is again based on the harmonic progression (2.4). More precisely, according to the model prediction, the ratios of these total amounts $S_2/S_3$, $S_2/S_4$, $S_2/S_5$, ... should be correspondingly equal to the ratios of the second member 1/2 of the harmonic progression (2.4) to its subsequent members 1/3, 1/4, 1/5, ... that is to values 3/2, 4/2, 5/2, ....

Fig. 14.3 presents the confirmation of this model prediction by the comparison of the amount $S_2$ of each of 16 doublets to the total amounts $S_3$, $S_4$, $S_5$ of $n$-plets ($n = 3, 4, 5$), which start with this doublet, in the human chromosome №1.

| DOUBLETS | TRIPLETS | TETRAPLETS | PENTAPLETS | $S_2/S_3$ | $S_2/S_4$ | $S_2/S_5$ |
|----------|----------|------------|------------|-----------|-----------|-----------|
| $S_2 = \Sigma(\text{AA})$ | $S_1 = \Sigma(\text{AAN})_k$ | $S_1 = \Sigma(\text{AANN})_{16}$ | $S_1 = \Sigma(\text{AANNN})_{64}$ | 1.50 | 2.00 | 2.50 |
| 10952057 | 7300224 | 5476855 | 4381298 | 1.50 | 2.00 | 2.50 |
| $S_2 = \Sigma(\text{AT})$ | $S_1 = \Sigma(\text{ATN})_k$ | $S_1 = \Sigma(\text{ATNN})_{16}$ | $S_1 = \Sigma(\text{ATNANN})_{64}$ | 1.50 | 2.00 | 2.50 |
| 8561194 | 5706906 | 4280647 | 3420561 | 1.50 | 2.00 | 2.50 |
| $S_2 = \Sigma(\text{AC})$ | $S_1 = \Sigma(\text{ACN})_k$ | $S_1 = \Sigma(\text{ACNN})_{16}$ | $S_1 = \Sigma(\text{ACNANN})_{64}$ | 1.50 | 2.00 | 2.50 |
| 5799729 | 3868541 | 2899991 | 2322063 | 1.50 | 2.00 | 2.50 |
| $S_2 = \Sigma(\text{AG})$ | $S_1 = \Sigma(\text{AGN})_k$ | $S_1 = \Sigma(\text{AGNN})_{16}$ | $S_1 = \Sigma(\text{AGNANN})_{64}$ | 1.50 | 2.00 | 2.50 |
| 8224510 | 5484720 | 4111320 | 3289579 | 1.50 | 2.00 | 2.50 |
| $S_2 = \Sigma(\text{TA})$ | $S_1 = \Sigma(\text{TAN})_k$ | $S_1 = \Sigma(\text{TANN})_{16}$ | $S_1 = \Sigma(\text{TANANN})_{64}$ | 1.50 | 2.00 | 2.50 |
| 7274275 | 4849731 | 3636741 | 2909412 | 1.50 | 2.00 | 2.50 |
| $S_2 = \Sigma(\text{TT})$ | $S_1 = \Sigma(\text{TTN})_k$ | $S_1 = \Sigma(\text{TTNN})_{16}$ | $S_1 = \Sigma(\text{TTNANN})_{64}$ | 1.50 | 2.00 | 2.50 |
| 11026157 | 7346507 | 551908 | 4409900 | 1.50 | 2.00 | 2.50 |
| $S_2 = \Sigma(\text{TC})$ | $S_1 = \Sigma(\text{TCN})_k$ | $S_1 = \Sigma(\text{TCNN})_{16}$ | $S_1 = \Sigma(\text{TCNANN})_{64}$ | 1.50 | 2.00 | 2.50 |
| 6923689 | 4617786 | 3461837 | 2768794 | 1.50 | 2.00 | 2.50 |
| $S_2 = \Sigma(\text{TG})$ | $S_1 = \Sigma(\text{TGN})_k$ | $S_1 = \Sigma(\text{TGNN})_{16}$ | $S_1 = \Sigma(\text{TGNANN})_{64}$ | 1.50 | 2.00 | 2.50 |
| 8396349 | 5598933 | 4198342 | 3357218 | 1.50 | 2.00 | 2.50 |
| $S_2 = \Sigma(\text{CA})$ | $S_1 = \Sigma(\text{CAN})_k$ | $S_1 = \Sigma(\text{CANN})_{16}$ | $S_1 = \Sigma(\text{CANNNN})_{64}$ | 1.50 | 2.00 | 2.50 |
| 8382478 | 5591208 | 4191829 | 3354600 | 1.50 | 2.00 | 2.50 |
| $S_2 = \Sigma(\text{CT})$ | $S_1 = \Sigma(\text{CTN})_k$ | $S_1 = \Sigma(\text{CTNN})_{16}$ | $S_1 = \Sigma(\text{CTNANN})_{64}$ | 1.50 | 2.00 | 2.50 |
| 8221421 | 5477836 | 4111963 | 3289510 | 1.50 | 2.00 | 2.50 |
| $S_2 = \Sigma(\text{CC})$ | $S_1 = \Sigma(\text{CNC})_k$ | $S_1 = \Sigma(\text{CCNN})_{16}$ | $S_1 = \Sigma(\text{CCNANN})_{64}$ | 1.50 | 2.00 | 2.50 |
| 6233384 | 4153642 | 3117570 | 2492824 | 1.50 | 2.00 | 2.50 |
| $S_2 = \Sigma(\text{CG})$ | $S_1 = \Sigma(\text{CGN})_k$ | $S_1 = \Sigma(\text{CGNN})_{16}$ | $S_1 = \Sigma(\text{CGNANN})_{64}$ | 1.50 | 2.00 | 2.50 |
| 1187593 | 789995 | 592235 | 475262 | 1.50 | 2.00 | 2.50 |
| $S_2 = \Sigma(\text{GA})$ | $S_1 = \Sigma(\text{GAN})_k$ | $S_1 = \Sigma(\text{GANN})_{16}$ | $S_1 = \Sigma(\text{GANNN})_{64}$ | 1.50 | 2.00 | 2.50 |
| 6923938 | 4612792 | 3462012 | 2768171 | 1.50 | 2.00 | 2.50 |
Fig. 14.3. The comparison of total amounts $S_2 = \Sigma (NN)$ of each of 16 doublets NN to the total amounts $S_3$ of 4 triplets, $S_4$ of 16 tetraplets, and $S_5$ of 64 pentaplets, which start with such attributive doublet NN, is shown for the human chromosome $N_1$. The left part of the table indicates the values of the corresponding total amounts. The right part contains appropriate values of the ratios $S_2/S_3$, $S_2/S_4$, and $S_2/S_5$, which are equal to the same magnitudes 1.5, 2.0, and 2.5 for the cases of all 16 doublets. Here N refers to any of nucleotides A, T, C, and G.

The rows in the left part of Fig. 14.3 shows very different numeric series of total amounts, which are individual in each of rows. But the right part shows that in each row its amounts are interrelated identically based on the numeric series of the ratios 1.5, 2.0, and 2.5, which serves here as a general invariant for the cases of all 16 doublets. But this sequence of ratios exists in the harmonic progression (2.4): 1, 1/2, 1/3, 1/4, 1/5, …, where the ratios of its second member 1/2 to its third, fourth and fifth members (that is, 1/3, 1/4, and 1/5) give this series 3/2, 4/2, and 5/2. Similar results are true for all other human chromosomes and for all those genomes, which were analyzed by the author.

The model predicts similarly the following numeric interconnections in the complete genomic sequences:

- The amount $S_3$ of any of 64 triplets NNN is algebra-harmonically interrelated with the total amounts $S_4$, $S_5$, $S_6$, … of oligomers in the following sets: 64 tetraplets, which start with this attributive triplet NNN; 256 pentaplets, which start with this attributive triplet NNN; 1024 six-plets, which start with this attributive triplet NNN;… The ratios of these total amounts $S_3/S_4$, $S_3/S_5$, $S_3/S_6$, … should be correspondingly equal to the ratios of the third member 1/3 of the harmonic progression (2.4) to its subsequent members 1/4, 1/5, 1/6, …, that is to values 4/3, 5/3, 6/3, …
- The amount $S_4$ of any of 256 tetraplets NNNN is algebra-harmonically interrelated with the total amounts of $S_5$, $S_6$, $S_7$, … of oligomers in the following sets: 256 pentaplets, which start with this attributive tetraplet NNNN; 1024 six-plets, which start with this attributive tetraplets NNNNN; 4906 seven-plets, which start with this attributive tetraplets NNNNN,…. The ratios of these total amounts $S_4/S_5$, $S_4/S_6$, $S_4/S_7$,… should be correspondingly equal to the ratios of the fourth member 1/4 of the harmonic progression (2.4) to its subsequent members 1/5, 1/6, 1/7, …, that is to values 5/4, 6/4, 7/4, …
- And so on (the length of attributive oligomers NN…N in the considered sets of n-plets should not be too large compared to the full length of the genomic sequence).

Similar model predictions exist not only for the listed cases, when the considered attributive nucleotides, or attributive doublets, or attributive triplets, etc. occupy the first positions in n-plets of the considered sets, but also for cases when
these attributive nucleotides or oligomers occupy there the second positions, or the third positions, etc (see corresponding rules about collective probabilities in oligomer tetra-groups for cases of locations of attributive oligomers in different positions of $n$-plets in the article [Petoukhov, 2018b].

Most of the long list of predictions, stemming from this quantum information model, is still awaiting their checking through analysis of various genomes. So far, the author has conducted only a relatively small number of checks of such predictions and has not found a single case of a phenomenological refutation of these predictions. The author will be grateful to those members of the scientific community who will find in the full-length sequences of different genomes such cases where these model predictions are not fulfilled.

These and other confirmed predictions of the model enlarge significantly the list of hyperbolic rules in genomes and lead to new tools and opportunities to study genetic structures. The obtained phenomenological data and the set of confirmed predictions of the quantum-information model testify that the eukaryotic and prokaryotic genomes represent a regular algebraic multi-level fractal-like net with important participation of the harmonic progression (2.4) in interconnections of its parts. This allows us to say about the algebraic harmony in living bodies. In theoretical biology, the quantum-information model has appeared, which allows one to predict with high accuracy a large number of quantitative interconnections between different kinds and sets of oligomers in eukaryotic and prokaryotic genomes.

15. Some concluding remarks

As is known, mutations and the pressure of natural selection influence the genomic sequences of nucleotides. For these reasons, one can assume that as a result of many millions of years of biological evolution, genomic sequences, due to various influences, receive a completely random structure as a whole. This article provides evidence that, despite mutations, the pressure of natural selection, and other evolutionary factors, the nucleotide sequences of the eukaryotic and prokaryotic genomes have universal algebraic invariants. One can believe that the algebraic unity of living organisms is found (this should be tested further and further on more and more number of genomes). New mathematical tools and approaches for an in-depth study of this world and its evolution appear.

The discovery of the described genomic invariants gives new knowledge about the unity of the world of all living organisms and about the features of biological evolution. This concerns additionally the problem of the origin of life, since the following natural question arises: where and how did these genomic algebraic invariants come from, which are expressed in the described hyperbolic (harmonic) rules and related to the quantum-information model if they exist even in the genomes of archaea and bacteria?

The genomic invariants, described in the article, are connected with hyperbolic sequences and transformations of hyperbolic rotations that shift the hyperbolic sequence along with itself. Hyperbolic rotations, which are also called Lorentz transformations and known in the special theory of relativity, draw attention to the structural connection of genetic phenomena with the hyperbolic geometry of the Minkowski plane. One of the well-known models of two-dimensional hyperbolic geometry is the Poincaré disk model, also called the conformal disk model. The Poincaré disk model is connected with split-quaternions by J. Cockle and seems to be interesting for studying some genetic structures and inherited physiological
phenomena as it was mentioned in previous author's publications on matrix genetics (see, for example, [Petoukhov, 2012]).

Living organisms are informational entities in which everything is subordinate to the task of reliably transmitting genetic information to descendants. All inherited physiological systems as parts of a whole organism must be structurally coupled with a genetic code for transmission to descendants in encoded form. Therefore, inherited physiological macrostructures can bear the imprint of structural features of the genetic code. For this reason, structural analogies exist between the genetic system and the properties of inherited physiological systems, for example, the unified properties of different sensor systems, which are reflected in the main psychophysical Weber-Fechner law [Petoukhov, 2016, 2019b, 2020a-c].

The question on a possible deep connection of physiology and brain functioning with principles of quantum informatics is considered in publications on many authors [Abbott, Davies, Pati, 2008; Altaisky, Filatov, 2001; Fimmel, Petoukhov, 2020; Igamberdiev, 1993, 2004; Matsuno, Paton, 2000; Patel, 2001a-c; Penrose, 1996; Petoukhov, 2018a, 2019b]. The results presented in this article give new essential materials to this perspective direction of thoughts.

It should be noted that the genomic hyperbolic rules are cardinaly different from well-known hyperbolic Zipf's law. Zipf's law was originally formulated in terms of quantitative linguistics, stating that given some corpus of natural language utterances, the frequency of any word is inversely proportional to its rank in the frequency table (see, for example, [Fagan, Gençay, 2010]). In linguistics and in other fields, Zipf's law speaks on the frequency of encounter of separate words or other separate objects. In contrast, the hyperbolic rules of the genomes focus on OS-sequences of the total amounts of n-plets and the genomic tetra-entanglement, that is, on the relative number of not separate oligomers, but the whole sums of sets of different n-plets distributed inside the genomic sequence, where each separate nucleotide is a part of many oligomers set existing simultaneously (each nucleotide is a distributed participant of many members of the appropriate genomic OS-sequence at once and makes a contribution to each of them). From the point of view of the quantum-information model, OS-sequences serve as quantum-information characteristics of genomic sequences.

The proposed oligomer sums method and the quantum-information model give new opportunities to study genetic systems and the inherited algebra-harmonic organization of living bodies. 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 & Burks, 1989]. Kepler did not make his own astronomic observations, but he found – in the huge astronomic data of Tycho Brahe - his Kepler’s laws of symmetric movements of planets relative to the Sun along ellipses. The author is convinced that further studies of symmetries in genetic and other physiological structures will reveal many more wonderful secrets of living matter.

The presented study is a continuation of the author's researches on symmetries in biological objects described in his publications (see References below). This study further illustrates the effectiveness of symmetry analysis in natural systems. No wonder the theory of symmetries is one of the foundations of modern mathematical
natural science. The presented results reveal the existence of a new broad class of symmetries in eukaryotic and prokaryotic genomes. They are connected with previous rules of a generalized symmetry for collective probabilities of sub-alphabets of $n$-plets in long DNA sequences, which were described by the author in the article [Petoukhov, 2018b] and whose importance were noted in the article “Petoukhov’s rules on symmetries in long DNA-texts” [Darvas, 2018]. In this article, the head of the International Institute “Symmetrion” (Budapest, Hungary) proposed to launch a corresponding international project: “Now, Petoukhov’s above rules of symmetries are candidates for the role of universal rules of long DNA-texts in living bodies. Further researches are needed to determine the degree of universality of these rules. Taking into account the huge number of species and long DNA-texts to be tested in these relations, I propose to launch an international project to study these genetic symmetries. Symmetrion initiates and can take part as a center of such an international project” [Darvas, 2018].

Appendix I. Numeric data on some epi-chains of the human chromosome № 1.

This Appendix shows numeric data about epi-chains represented graphically above in Figs. 13.2.—13.6.

| n  | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   |
|----|------|------|------|------|------|------|------|------|------|------|
| A  | Real | 35337501 | 16768845 | 11179286 | 8383461 | 6706672 | 5588773 | 4792078 | 4912017 | 3726860 | 3354107 |
|    | Model| 35337501 | 16768751 | 11179167 | 8384375 | 6707500 | 5589584 | 4791072 | 4912188 | 3726389 | 3353750 |
| Δ% | 0    | -0.001 | -0.001 | 0.011 | 0.012 | 0.015 | -0.021 | 0.004 | -0.003 | -0.011 |
| T  | Real | 33620498 | 16808862 | 11207274 | 8405040 | 6724359 | 5601854 | 4801395 | 4202773 | 3735327 | 3360459 |
|    | Model| 33620498 | 16810249 | 11206833 | 8405125 | 6724100 | 5603416 | 4802928 | 4202562 | 3735611 | 3362050 |
| Δ% | 0    | 0.008  | -0.004 | 0.001 | -0.004 | 0.028  | 0.032  | -0.005 | 0.008  | 0.047  |
| C  | Real | 24024903 | 12013624 | 8005708  | 6008215  | 4803919  | 4002753  | 3433636  | 3005311  | 2668499  | 2402186  |
|    | Model| 24024903 | 12012451 | 8008301  | 6006226  | 4804981  | 4004151  | 3432192  | 3003113  | 2669434  | 2402490  |
| Δ% | 0    | -0.010 | 0.032  | -0.033 | 0.022   | 0.035   | -0.044  | 0.013   | 0.035   | 0.013   |
| G  | Real | 24057606 | 12028924 | 8021235  | 6013412  | 4813156  | 4013372  | 3435824  | 3006763  | 2673815  | 2407301  |
|    | Model| 24057606 | 12028803 | 8019202  | 6014402  | 4811521  | 4009601  | 3436801  | 3007201  | 2673067  | 2405761  |
| Δ% | 0    | -0.001 | -0.025 | 0.016   | -0.034  | -0.094  | 0.028   | 0.015   | -0.028  | -0.064  |

| n  | 11   | 12   | 13   | 14   | 15   | 16   | 17   | 18   | 19   | 20   |
|----|------|------|------|------|------|------|------|------|------|------|
| A  | Real | 3048864 | 2794792 | 2579808 | 2355536 | 2235833 | 2106904 | 1972794 | 1863195 | 1767938 |
|    | Model| 3048864 | 2794792 | 2579808 | 2355536 | 2235833 | 2106904 | 1972794 | 1863195 | 1767938 |
| Δ% | -0.021 | 0.055  | 0.015  | 0.038  | 0.000  | 0.010  | -0.004 | 0.001  | -0.056  | -0.063  |
| T  | Real | 3054627 | 2802390 | 2584949 | 2400749 | 2240133 | 2101692 | 1976639 | 1866079 | 1768246 | 1680401 |
|    | Model| 3056409 | 2801708 | 2586192 | 2401464 | 2241367 | 2102817 | 1977676 | 1867805 | 1769500 | 1681025 |
| Δ% | 0.058  | -0.024 | -0.089 | 0.030  | 0.055  | -0.020 | 0.052  | 0.092  | 0.071  | 0.037  |
| C  | Real | 2185430 | 2001255 | 1845496 | 1781876 | 1600752 | 1501210 | 1413849 | 1335576 | 1263741 | 1200991 |
|    | Model| 2184082 | 2002075 | 1848069 | 1710665 | 1610660 | 1505556 | 1413230 | 1334717 | 1264469 | 1201245 |
| Δ% | -0.063 | 0.041  | 0.139  | -0.152 | 0.057  | 0.023  | -0.051 | -0.064 | 0.058  | 0.021  |
| G  | Real | 2186817 | 2006465 | 1851228 | 1717404 | 1605990 | 1503735 | 1414212 | 1337414 | 1267181 | 1202700 |
|    | Model| 2187055 | 2004801 | 1850585 | 1718400 | 1603840 | 1503600 | 1415155 | 1336534 | 1260190 | 1202880 |
| Δ% | 0.011  | -0.083 | -0.035 | 0.050  | -0.134 | -0.009 | 0.067  | -0.066 | -0.078 | 0.015  |
**Fig. 1.1.** The results of analysis - by the oligomer sums method – the nucleotide sequence of the epi-chain of the second order N_{2/1} (Fig. 13.1b), which consists of nucleotides with serial numerations 1-3-5-7-9-… in the DNA sequence of the human chromosome № 1. The table demonstrates that the model hyperbolic progressions S_A/n, S_T/n, S_C/n, S_G/n (in red) almost completely coincide with the OS-sequences of real total amounts of those n-plets, which start with a nucleotide A, or T, or C, or G in this epi-chain correspondingly. Differences between the corresponding values in these numerical sequences are expressed by shown small percentage values Δ%.

| n | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| A Real | 22360413 | 11179286 | 7453552 | 5588773 | 4472245 | 3726860 | 3196917 | 2792365 | 2483348 | 2176711 |
| Model | 22360413 | 11180207 | 7453471 | 5590103 | 4472083 | 3726736 | 3194345 | 2789052 | 2484490 | 2179243 |
| Δ% Δ% | 0.008 | -0.001 | 0.024 | -0.004 | -0.003 | -0.081 | 0.064 | 0.046 | 0.052 | 0.052 |

**Fig. 1.2.** The results of analysis - by the oligomer sums method – the nucleotide sequence of the epi-chain of the third order N_{3/1} (Fig. 13.1d), which consists of nucleotides with serial numerations 1-4-7-10-13-… in the DNA sequence of the human chromosome № 1. The table demonstrates that the model hyperbolic progressions S_A/n, S_T/n, S_C/n, S_G/n (in red) almost completely coincide with the OS-sequences of real total amounts of those n-plets, which start with a nucleotide A,
or T, or C, or G in this epi-chain correspondingly. Differences between the corresponding values in these numerical sequences are expressed by shown small percentage values $\Delta\%$.

| n | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   |
|---|------|------|------|------|------|------|------|------|------|------|
| A | Real | 16768845 | 8383461 | 5588773 | 4192017 | 3354107 | 2793265 | 2394635 | 2095893 | 1863181 | 1677938 |
| Model | 16768845 | 8384423 | 5598615 | 4192211.25 | 3353769 | 2794808 | 2395549 | 2096106 | 1863205 | 1676885 |
| $\Delta\%$ | 0 | 0.011 | 0.015 | 0.005 | -0.010 | 0.055 | 0.038 | 0.010 | 0.001 | -0.063 |
| T  | Real | 16808862 | 8405040 | 5601854 | 4202773 | 3366459 | 2802390 | 2400749 | 2101692 | 1866079 | 1680401 |
| Model | 16808862 | 8404431 | 5602954 | 4202216 | 3367722 | 2801477 | 2401266 | 2101086 | 1867651 | 1680886 |
| $\Delta\%$ | 0 | -0.007 | 0.020 | -0.013 | 0.039 | -0.033 | 0.022 | -0.028 | 0.084 | 0.029 |
| C  | Real | 12013624 | 6008215 | 4002753 | 3003511 | 2402186 | 2001255 | 1718676 | 1501210 | 1335576 | 1200991 |
| Model | 12013624 | 6006312 | 4001544 | 3003406 | 2402725 | 2002271 | 1716332 | 1501703 | 1334847 | 1201362 |
| $\Delta\%$ | 0 | -0.023 | 0.045 | -0.003 | 0.022 | 0.051 | -0.142 | 0.033 | -0.055 | 0.031 |
| G  | Real | 12028924 | 6013412 | 4013372 | 3006763 | 2407301 | 2006465 | 1717404 | 1503735 | 1337414 | 1202700 |
| Model | 12028924 | 6014462 | 4009641 | 3007231 | 2405785 | 2004821 | 1718418 | 1503616 | 1336547 | 1202892 |
| $\Delta\%$ | 0 | 0.017 | -0.093 | 0.016 | -0.063 | -0.082 | 0.059 | -0.008 | -0.065 | 0.016 |

Fig. I.3. The results of analysis - by the oligomer sums method – the nucleotide sequence of the epi-chain of the 4th order $\text{N}_4/1$, which consists of nucleotides with serial numerations 1-5-9-13-... in the DNA sequence of the human chromosome $\text{N}_1$. The table demonstrates that the model hyperbolic progressions $S_A/n$, $S_T/n$, $S_C/n$, $S_G/n$ (in red) almost completely coincide with the OS-sequences of real total amounts of those $n$-plets, which start with a nucleotide A, or T, or C, or G in this epi-chain correspondingly. Differences between the corresponding values in these numerical sequences are expressed by shown small percentage values $\Delta\%$. 


doi:10.20944/preprints202005.0471.v1
Table 1.1: Differences between the observed and calculated values of nucleotides in the DNA sequence of the human chromosome No 1.

| n   | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   |
|-----|------|------|------|------|------|------|------|------|------|------|
| A   | Real | 6706672 | 3354107 | 2235831 | 1677938 | 1341408 | 1117975 | 958626 | 839279 | 744475 | 670703 |
|     | Model| 3353336 | 2235557 | 1676668 | 1341334 | 1117779 | 958096 | 838334 | 745186 | 670667 |
| Δ%  | 0    | -0.023 | -0.012 | -0.076 | -0.005 | -0.018 | -0.055 | -0.113 | 0.095  | -0.005 |
| T   | Real | 6724359 | 3360459 | 2240133 | 1680401 | 1344421 | 1119290 | 961102 | 839090 | 746575 | 672348 |
|     | Model| 3362179 | 2241455 | 1681090 | 1344872 | 1120727 | 960623 | 840545 | 747131 | 672436 |
| Δ%  | 0    | -0.051 | -0.039 | -0.076 | -0.034 | -0.128 | -0.079 | 0.088  | 0.077  | 0.013  |
| C   | Real | 4803919 | 2402186 | 1600752 | 1200991 | 961518 | 800223 | 686222 | 601012 | 533486 | 480738 |
|     | Model| 2401959 | 1601306 | 1200900 | 960764 | 800653 | 686274 | 600490 | 533769 | 480392 |
| Δ%  | 0    | -0.009 | -0.003 | -0.076 | -0.054 | -0.087 | -0.008 | 0.053  | -0.072 |
| G   | Real | 4813156 | 2407301 | 1605990 | 1202700 | 962275 | 803863 | 686639 | 609918 | 536368 | 481023 |
|     | Model| 2403156 | 1604585 | 1203289 | 962631 | 802193 | 687594 | 601645 | 534795 | 481316 |
| Δ%  | 0    | -0.030 | -0.100 | 0.049  | 0.037  | -0.208 | 0.139  | 0.121  | -0.294 |

Fig. 1.4. The results of analysis - by the oligomer sums method – the nucleotide sequence of the epi-chain of the 10th order N_{10/1}, which consists of nucleotides with serial numerations 1-11-21-31-41-… in the DNA sequence of the human chromosome No 1. The table demonstrates that the model hyperbolic progressions S_n/n, S_T/n, S_C/n, S_G/n (in red) almost completely coincide with the OS-sequences of real total amounts of those n-plots, which start with a nucleotide A, or T, or C, or G in this epi-chain correspondingly. Differences between the corresponding values in these numerical sequences are expressed by shown small percentage values Δ%.
| Epi-ch. | S_A/(S_A+S_T+S_C+S_G) | S_T/(S_A+S_T+S_C+S_G) | S_C/(S_A+S_T+S_C+S_G) | S_G/(S_A+S_T+S_C+S_G) |
|--------|------------------------|------------------------|------------------------|------------------------|
| N1/1   | 0.2910                 | 0.2918                 | 0.2085                 | 0.2087                 |
| N2/1   | 0.2910                 | 0.2917                 | 0.2085                 | 0.2088                 |
| N3/1   | 0.2910                 | 0.2917                 | 0.2084                 | 0.2088                 |
| N4/1   | 0.2910                 | 0.2917                 | 0.2085                 | 0.2088                 |
| N10/1  | 0.2910                 | 0.2918                 | 0.2084                 | 0.2088                 |
| N50/1  | 0.2910                 | 0.2917                 | 0.2086                 | 0.2088                 |
Acknowledgments

Some results of this paper have been possible due to a 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 selforganizing systems”, where the author 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, I.V. Stepanyan, V.I. Svirin and G.K. Tolokonnikov for their collaboration. The author is also grateful to the members of the International Symmetry Association (Budapest, http://isa.symmetry.hu/) and the International Seminar "Algebraic Biology and System Theory" (Moscow, https://www.youtube.com/channel/UC8JLsuRzzPsRiHwrwEjMctw) for discussing the author's researches in the field of matrix genetics and algebraic biology.

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. Asymptotically increasing compliance of genomes with Chargaff’s second parity rules through inversions and inverted transpositions. Proc Natl Acad Sci USA, 103 (47), pp. 17828–17833 (2006), doi:10.1073/pnas.0605553103

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

Amico L., Fazio R., Osterloh A., Vedral V. Entanglement in many-body systems. Rev. Mod. Phys., 80, 517 (2008).

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

Fagan S., Gençay R. “An introduction to textual econometrics", in Ullah, Aman; Giles, David E. A. (eds.), Handbook of Empirical Economics and Finance, CRC Press, pp. 133–153, (2010), ISBN 9781420070361

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

Fimmel E., Gumbel M., Karpuzoglu A., Petoukhov S. On comparing composition principles of long DNA sequences with those of random ones. Biosystems, v. 180, pp. 101-108, June 2019,

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 Artificial Systems for Medicine and Education III. AIMEE 2019. Advances in Intelligent Systems and Computing, vol 1126, p.126-135. Springer, Cham DOI https://doi.org/10.1007/978-3-030-39162-1_12

Frank-Kamenetskii M.D. The most important molecule. Moscow, Nauka (1988) (in Russian).

Graham R.L., Knuth D.E., Parashnik O. Concrete Mathematics. A Foundations for
Gühne O., Tóth G. Entanglement detection. *Phys. Rep.*, 474, pp. 1-75 (2009).

Harkin A.A., Harkin J.B. Geometry of Generalized Complex Numbers. *Mathematics Magazine*, v. 77(2), p. 118-29 (2004).

Horodecki R., Horodecki P., Horodecki M., Horodecki K. Quantum entanglement. *Rev. Mod. Phys.*, 81, 865 (2009);

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

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

Igamberdiev A.I. Quantum computation, non-demolition measurements, and reflective control in living systems. – *BioSystems*, v. 77, pp. 47–56 (2004).

Kantor I.L., Solodovnikov A.S. Hypercomplex numbers. Berlin, New York: Springer-Verlag (1989). ISBN 978-0-387-96980-0.

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

McFadden J., Al-Khalili J. The origins of quantum biology. *Proceedings of the Royal Society A*, Vol. 474, Issue 2220, p. 1-13, 12 December 2018, https://doi.org/10.1098/rspa.2018.0674

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

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

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

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

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

Petoukhov S.V. Matrix genetics and algebraic properties of the multi-level system of genetic alphabets. *Neuroquantology*, v. 9, №4, p. 60-81 (2011).

Petoukhov S.V. Symmetries of the genetic code, hypercomplex numbers and genetic matrices with internal complementarities. *Symmetry: Culture and Science*, 2012, vol. 23, № 3-4, p. 275-301. http://petoukhov.com/PETOUKHOV_GENETIC_MATRICES_COMPLEMENTARITIES.pdf

Petoukhov S.V. The system-resonance approach in modeling genetic structures. *Biosystems*, v. 139, p. 1-11 (January 2016).

Petoukhov S.V. Genetic coding and united-hypercomplex systems in the models of algebraic biology. *Biosystems*, v. 158, August 2017, pp. 31–46 (2017).

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

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

Patel A. Quantum algorithms and the genetic code. *Pramana – Journal of Physics*, v. 56, 2-3, pp. 367-381 (2001a), arXiv:quant-ph/0002037
Petoukhov S.V. The rules of long DNA-sequences and tetra-groups of oligonucleotides. arXiv:1709.04943v5, 5th version from 8 October 2018, 159 pages (2018b).

Petoukhov S.V. Nucleotide Epi-Chains and New Nucleotide Probability Rules in Long DNA Sequences. Preprints 2019, 2019040011 (doi:10.20944/preprints201904.0011.v2), (2019a).

Petoukhov S.V. 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, v. 938. Springer, Cham. Online publication on 29 March 2019. (2019b).

Petoukhov S.V. Hyperbolic Numbers in Modeling Genetic Phenomena. Preprints 2019, 2019080284, 36 p. (2020a), doi:10.20944/preprints201908.0284.v4.

Petoukhov S.V. The Genetic Code, Algebraic Codes and Double Numbers. Preprints 2019, 2019110301 (2020b) (doi:10.20944/preprints201911.0301.v2).

Petoukhov S.V. Hyperbolic Numbers, Genetics and Musicology. In: Hu Z., Petoukhov S., He M. (eds) Advances in Artificial Systems for Medicine and Education III. AIMEE 2019. Advances in Intelligent Systems and Computing, vol. 1126, p. 195-207. Springer, Cham (2020c). DOI https://doi.org/10.1007/978-3-030-39162-1_18

Petoukhov S.V., He M. Symmetrical Analysis Techniques for Genetic Systems and Bioinformatics: Advanced Patterns and Applications. IGI Global, USA (2010). http://petoukhov.com/Petoukhov%20He%20-%202010-%20Symmetrical%20Analysis%20Techniques%20for%20Genetic%20Systems%20and%20Bioinformatics.pdf.

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

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

Petoukhov S.V., Petukhova E.S., Svirin V.I. Symmetries of DNA alphabets and quantum informational formalisms. Symmetry: Culture and Science, v. 30, No. 2, pp. 161-179 (2019), https://doi.org/10.26830/symmetry_2019_2_161, http://petoukhov.com/PETOUKHOV%20GENETIC%20QUANTUM%20INFORMATIONAL%20MODEL%202019.pdf

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

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

Rosandic M., Vlahovic I., Gluncic M., and Paar V. Trinucleotide’s quadruplet symmetries and natural symmetry law of DNA creation ensuing Chargaff’s second
parity rule. *Journal of Biomolecular Structure and Dynamics*, 34:7, pp.1383-1394, (2016), DOI: 10.1080/07391102.2015.1080628

**Shporer S., Chor B., Rosset S., and Horn D.** Inversion symmetry of DNA k-mer counts: validity and deviations. *BMC Genomics*, 17(1): 696 (2016).

**Walter M., Gross D., Eisert J.** Multi-partite entanglement. (2017). arXiv:1612.02437

**Yamagishi M.E.B.** *Mathematical Grammar of Biology*. Switzerland, Springer International Publishing AG (2017). ISBN 978-3-319-62689-5.