Block-Stochastic organization of genomes, stochastics in biomechanics of motions, and inherited Gestalt principles in physiology

Cite as: AIP Conference Proceedings **2697**, 060004 (2023); https://doi.org/10.1063/5.0111815
Published Online: 16 March 2023

V. V. Verevkin and S. V. Petoukhov

Block-Stochastic Organization of Genomes, Stochastics in Biomechanics of Motions, and Inherited Gestalt Principles in Physiology

V.V. Verevkin and S.V. Petoukhov[a)]

Mechanical Engineering Research Institute of the Russian Academy of Sciences, 4, M. Kharitonyevskiy Lane, Moscow, 101990, Russia

[a)] *Corresponding author: spetoukhov@gmail.com*

Abstract. To reveal the hidden regularities of the stochastic organization of nucleotide sequences in informational single-stranded DNA molecules of genomes of higher and lower biological species, the following author's method was applied: each such sequence is represented as a set of parallel n-texts, each of which is written in its own n-plets alphabet of DNA (among these alphabets include alphabets of 4 nucleotides, 16 duplets, 64 triplets, etc.). A comparative analysis of the percentage compositions of n-plets in these n-texts is carried out and the universal regularities of the multilevel block-stochastic organization of genomic DNA information sequences are revealed. The analogies of the phenomena of block-stochastic organization of these DNAs with genetically inherited physiological phenomena of similar types are discussed, including the Bernstein problem in the field of biomechanics of movements and also the ability of the brain, studied by Gestalt psychology.

INTRODUCTION

Advances in molecular biology have led to a new understanding of life itself: "*Life is a partnership between genes and mathematics*" [1]. Science reveals more and more the extent to which genetic inheritance determines the characteristics of a person's life, including his preferences for types of nutrition, behavior, emotional mood, perception, etc. All physiological structures of the body for transmission to descendants must be genetically encoded, and therefore they carry on themselves the imprints of the structural features of the genetic coding system [2]. Genetics as a science began with the discovery by G. Mendel of stochastic rules for the inheritance of traits in experiments on the crossing of organisms. The study of the regularities of stochastic phenomena of genetic inheritance at different levels of living matter continues in modern science.

Many processes in living bodies are of a stochastic nature and proceed against a background of noise or are accompanied by noise. For example, the expressions"gene noise" or "cell noise", which are known in biology, reflect the fact that even genetically identical cells within the same tissue exhibit different levels of protein expression, different sizes and structures due to the stochastic nature of interactions of individual molecules in cells. This stochastic nature of genetic inheritance is manifested, in particular, in the fact that all people, even identical twins, have different fingerprints. In general, living bodies can be viewed as a certain phenomenon of a kind of block-stochastic organization, the structural features of which are subject to study. These features include genetically inherited multiblock structures, in which individual blocks as a whole (globally) are similar to each other, although locally they differ significantly (like fingers in humans).

In the laboratory for research of biomechanical systems IMASH, planned work is underway to develop genetic biomechanics, which studies the structural relationships of the genetic coding system with the features of inherited biomechanical systems [2-5]. The novelty of this scientific direction can be judged by the fact that in a large number of monographs and textbooks on biomechanics, until very recently, no connection between biomechanical structures

Proceedings of International Conference of Young Scientists and Students "Topical Problems of Mechanical Engineering" ToPME 2021
AIP Conf. Proc. 2697, 060004-1–060004-6; https://doi.org/10.1063/5.0111815
Published by AIP Publishing. 978-0-7354-4416-4/$30.00

with structures of genetic coding was noted at all [6-12]. Many of the laboratory's works on genetic biomechanics are publicly available on the website http://petoukhov.com/.

The aim of this study is to analyze the numerical symmetries and rules of the block-stochastic organization of DNA nucleotide sequences in the genomes of higher and lower organisms, as well as the development of model concepts of the inherited mechanisms of the dictatorial influence of molecules DNA on biological macrostructures. Among the studied phenomena are genetically inherited phenomena of biomechanical movements described by the classic of biomechanics N.A. Bernstein, as well as the phenomena of brain activity studied by Gestalt psychology.

MATERIALS AND RESEARCH METHODS

In this work, the initial data on the nucleotide sequences of single DNA strands of different genomes were taken from the publicly available GenBank genomic data bank (https://www.ncbi.nlm.nih.gov/genbank/). These genomic DNA sequences consist of many millions of nucleotides (molecular "letters") of four types: adenine A, guanine G, cytosine C, and thymine T. As it is known, DNA contains alphabets of 4 nucleotides, 16 duplets (i.e. combinations of nucleotides two each, for example, AA, AG, AC, ...), 64 triplets (combinations of nucleotides of three, for example, AAA, AAC, AAG, ...), etc.

A feature of our research approach to the analysis of genomes, which turned out to be effective, is the consideration of each of the genomic nucleotide sequences of a single DNA strand as a set of parallel n-texts, each of which is written in its own n-alphabet. For example, the nucleotide sequence CAGGATCGACGT ... is represented as a set of n-texts: namely as 1-text C-A-G-G-A-T-C-G-A-C-G-T- ..., 2-text CA-GG-AT-CG-AC-GT- ..., 3-text CAG-GAT-CGA-CGT- ..., 4-text CAGG-ATCG-ACGT-… and so on. In each of these n-textual representations of the studied genomic DNA sequence, the percentage of each of the n-plets kinds that are members of the corresponding n-alphabet is calculated. The resulting set of data on the percentage compositions of n-texts is analyzed to identify possible regular relationships of a stochastic nature in it.

In our previous publications on genetic biomechanics, the relationship of the DNA alphabet system of n-plets with the tensor family of square matrices was shown [2-5]. This connection, like other previously obtained results, testifies in favor of the possibility of modeling a genetic coding system based on the formalisms of quantum mechanics and quantum informatics, in which the tensor product plays an important role: the state space of a multi-component system is a tensor product of the state spaces of its components [13].

This article presents additional results on the use of tensor products in the analysis and modeling of single-stranded DNA nucleotide sequences in the genomes of higher and lower organisms to identify in them the numerical patterns of block stochastic organization.

RESULTS AND DISCUSSION

The results described, showing the existence of universal rules for the stochastic organization of genomes, at least for values n = 1, 2, 3, 4, 5, were obtained by analyzing the DNA of the following genomes:

1. all 24 human chromosomes;
2. all chromosomes of Drosophila, mouse, worm, many plants;
3. 19 genomes of bacteria and archaea;
4. -many extremophiles living in extreme conditions, including, for example, radiation with a level exceeding 1000 times fatal to humans.

Below these rules are illustrated by numerical data (Fig. 1) for n = 1, 2, 3 for the DNA of the first human chromosome containing a sequence of about 250 million nucleotides A, T, C, G.

The described analytical approach uses the vector [C, A, G, T], represented as the sum of four sparse vectors with only one nonzero coordinate (1):

$$\|C \quad A \quad G \quad T\| = \|C \quad 0 \quad 0 \quad 0\| + \|0 \quad A \quad 0 \quad 0\| + \|0 \quad 0 \quad G \quad 0\| + \|0 \quad 0 \quad 0 \quad T\| \qquad (1)$$

Similarly, each of the 4^n-dimensional vectors of the tensor family of vectors $[C, A, G, T]^{(n)}$ is represented as a sum of 4^n sparse vectors with only one nonzero coordinate. In this representation, we introduce short designations of the following type: [C] = [C,0,0,0]; [A] = [0,A,0,0]; [G] = [0,0,G,0]; [T] = [0,0,0,T]; [CC] = [CC,0,0,0,0,0,0,0,0,0, 0,0,0,0,0,0]; [AG] = [0,0,0,0,0,0,AG,0,0,0,0,0,0,0,0,0], etc. In this approach, 4 nucleotides C, A, T, G are represented by 4 vectors [C], [A], [T], [G]; 16 duplets CC, CA, ... are represented by 16

vectors [CC], [CA], ...; 64 triplets CCC, CCA, ... are represented by 64 vectors [CCC], [CCA], ...; etc. Each of these sparse vectors is referred to as a "parent". In addition, the tetra-vector [C, A, T, G] is also used, the components of which are symbols of 4 nucleotides. We consider tensor products (both on the right and on the left) of each of the parent vectors with this tetra-vector and its tensor powers $[C, A, T, G]^{(n)}$. Such tensor operations produce new character vectors of increased dimension, each of which is referred to as a "child vector" with respect to the corresponding parent vector. The components of the child vector are symbols of n-plets, each of which has a certain percentage of implementation in the corresponding n-text of the analyzed DNA.

%C	%T
%A	%G

=

0.2085	0.2910
0.2918	0.2087

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

=

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

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

=

0.01385	0.01878	0.01524	0.01861	0.01183	0.01977	0.01447	0.03693
0.01853	0.00291	0.01789	0.02104	0.01622	0.00254	0.02375	0.01988
0.01758	0.01275	0.00251	0.00227	0.01317	0.01942	0.01441	0.02237
0.02009	0.02088	0.00259	0.00291	0.02388	0.01781	0.01614	0.01848
0.01588	0.01964	0.01103	0.01986	0.01255	0.01456	0.00962	0.01960
0.02226	0.00233	0.01939	0.01284	0.01437	0.00253	0.01327	0.01756
0.01972	0.01981	0.01457	0.01947	0.00956	0.01115	0.01256	0.01600
0.03725	0.01884	0.01988	0.01895	0.01445	0.01534	0.01185	0.01382

FIGURE. 1. Matrices of the percentage composition of n-plets in n-texts of DNA in human chromosome # 1 (n = 1, 2, 3). The location of the percents symbols of n-plets in matrices corresponds to the location of n-plets in the tensor family of matrices $[C, A; T, G]^{(n)}$

Surprising and unexpected is the fact that in the chromosomal DNA under consideration, the sum of the individual percentages of all n-plets, which are components of the daughter vector, is, with high accuracy, equal to the percentage of the n-plet of the parent vector.

Let us illustrate this with numerical examples of the tensor product with the participation of the maternal vector [C], representing the cytosine C:

- [C] $\otimes$ [C, A, T, G] = [CC, CA, CT, CG] (this is the tensor product on the right). Let us denote the percentage sum %CC+%CA+%CT+%CG of the grouping of 4 duplets in the resulting 4-dimensional daughter vector by the symbol Σ%CN, since all these duplets contain nucleotide C in the first position (hereinafter, symbol N denotes any of the four nucleotides A, C , T, G). Using the phenomenological data on the percentage of these duplets from Fig. 1, we obtain the value of their sum in the duplet representation of this DNA: Σ%CN = 0.05409 + 0.07274 + 0.07134 + 0.01031 ≈ **0.2085**. But this sum is equal to the percentage of cytosine in the monoplet representation of this DNA in Fig. 1: %C = **0.2085**;

- [C, A, T, G] ⊗ [C] = [CC, AC, TC, GC] (this is the tensor product on the left). Let's denote the total percentage sum %CC+%AC+%TC+%GC of this new grouping of 4 duplets in the new 4-dimensional child vector by the symbol Σ%NC, since all of these doublets contain nucleotide C in the second position. Using the phenomenological data on the percentage of these 4 duplets from Fig. 1, we obtain the value of their sum in the duplet representation of this DNA: Σ%NC = 0.05409+0.05033+0.06008+0.04402 ≈ **0.2085**. This amount is also almost equal to the percentage of nucleotide C in the monoplet representation of this DNA: %C = **0.2085**;
- [C] ⊗ [C, A, T, G]$^{(2)}$ = [CCC, CCA, CCT, CCG, CAC, CAA, CAT, CAG, CTC, CTA, CTT, CTG, CGC, CGA, CGT, CGG]. Let us denote the total sum of percentages of 16 triplets in the resulting 16-dimensional daughter vector by the symbol Σ% CNN, since all these triplets contain nucleotide C in the first position. Using the values of individual percentages of triplets from Fig. 1, we obtain the value of the sum in the triplet representation of this DNA: Σ%CNN = 0.01385+0.01878+0.01853 +0.00291+0.01524 +0.01861+0.01789+0.02104+0.01758 +0.01275 +0.02009+0.02088+0.00251+0.00227+0.00259+0.00291 ≈ **0.20843**. This amount is also almost equal to the percentage of nucleotide C in the monoplet representation of this DNA: %C = **0.2085**;
- C, A, T, G] ⊗ [C] ⊗ [C, A, T, G] = [CC, AC, TC, GC] ⊗ [C, A, T, G] = [CCC, CCA, CCT, CCG, ACC, ACA, ACT, ACG, TCC, TCA, TCT, TCG, GCC, GCA, GCT, GCG]. Let us denote the total sum of the percentages of these new 16 triplets in the resulting new 16-dimensional child vector by the symbol Σ%NCN, since all these triplets contain nucleotide C in the second position. Using the values of individual percentages of triplets from Fig. 1, we obtain the value of the sum in the triplet representation of this DNA: Σ%NCN=0.01385+0.01878+0.01853+0.00291+0.01183+0.01977+0.01622+0.00254+0.01588+0.01964+0.0222+0.00233+0.01255+0.01456+0.01437+0.00253 = **0.2085**. This amount is also almost equal to the percentage of nucleotide C in the monoplet representation of this DNA: %C = **0.2085**;
- [C, A, T, G]$^{(2)}$ ⊗ [C] = [CCC, CAC, CTC, CGC, ACC, AAC, ATC, AGC, TCC, TAC, TTC, TGC, GCC, GAC, GTC, GGC]. Let us denote the total sum of percentages of all 16 triplets, which serve as coordinates in the newly formed 16-dimensional child vector, by the symbol Σ%NNC, since all these triplets contain nucleotide C in the third position. Using the values of individual percentages of triplets from Fig. 1, we obtain the value of the sum in the triplet representation of this DNA: Σ%NNC = 0.01385+0.01524+0.01758+0.00251+0.01183+0.01447+0.01317+0.01441+0.01588+0.01103+0.01972+0.01457+0.01255+0.00962+0.00956 +0.01256 ≈ **0.2085.** This amount is also almost equal to the percentage of nucleotide C in the monoplet representation of this DNA: %C = **0.2085**.

A similar equality to the value of %C is also true for the sums of percentages in the groupings of 64 tetraplets (in the tetraplet representation of this DNA) containing nucleotide C at the first, or second, or third, or fourth positions, that is, for the sums Σ%CNNN, Σ%NCNN, Σ%NNCN, Σ%NNNC, in which the sets of values of the summands are significantly different.

Similar equalities in n-plets representations of genomic DNA are also fulfilled in the case of the parent vectors of the remaining nucleotides [A], [T], [G] in their tensor product with the tetra-vector [C, A, T, G] (both on the right and left): the total sum of individual percentages of all n-plets, represented in the coordinates of the child vector, is, with high precision, equal to the nucleotide percentage in the parent vector.

Moreover, similar block-stochastic rules in genomic DNA are also fulfilled in the case of maternal vectors based on each of the 16 duplets ([CC], [CA], ...), each of 64 triplets ([CCC], [CCA], ...) , and so on, regarding the results of their tensor product with the tetra-vector [C, A, T, G] (both on the right and on the left). For example, consider the case of the maternal vector [CC] in the DNA of the same human chromosome # 1:

- [CC] ⊗[C, A, T, G] = [CCC, CCA, CCT, CCG]. Let us denote the sum of percentage sum %CCC+%CCA+%CCT+%CCG of these four triplets of the so generated 4-dimensional child vector by Σ%CCN, since all these triplets start with a duplet CC. Using the individual percentages of these four triplets in the triplet representation of DNA from Fig. 1, we get: Σ%CCN = 0.01385+0.01878+0.01853 +0.00291 ≈ **0.05407**. This sum is practically equal to %CC = **0.05409** in the duplet representation of this DNA in Fig. 1;
- [C, A, T, G] ⊗ [CC] = [CCC, ACC, TCC, GCC]. Let us denote the percentage sum %CCC +%ACC +%TCC +%GCC of this new grouping of 4 triplets of the so generated 4-dimensional child vector by the symbol Σ%NCC, since all these triplets end with a duplet CC. Using the individual percentages of these four triplets in the triplet representation of DNA from Fig. 1, we get:

Σ%NCC=0.01385+0.01183+0.01588+0.01255 ≈ **0.05411**. This sum is practically equal to %CC = **0.05409** in the duplet representation of this DNA in Fig. 1.

The set of similar results in our studies indicates a multilevel system of block-stochastic DNA organization in the genomes of higher and lower organisms.

The universal phenomena described above of the relative independence of the sums of the percentages of n-plets (in their block tetragroupings in the n-texts of genomic DNA) from the values of individual summands are analogs of gestalt images in psychology. Gestalt psychology studies the genetically inherited properties of the brain to form integral images, relatively independent of their particular components [13-16]. For example, a musical melody is recognized by us, even when it is played on different instruments and in different frequency ranges with a changed frequency composition of its components. The same applies to the visual perception of pictures. This is an inherited fundamental property of the psyche: to seek a whole in a disparate. Thanks to the ability to think in gestalts, you can understand the sentence, even if you change the order of the letters in each word and leave only the beginning and ending in place. For example, everyone can easily understand the following phrase, strongly "mutated" by local permutations: "Aoccdrnig to a rscheearch at Cmabrigde Uinervtisy, it deosn't mttaer in waht oredr the ltteers in a wrod are» (this "mutated" phrase example is taken from https://www.dictionary.com/e/typoglycemia/).

The analogies between Gestalt perception phenomena and the phenomena in the genomic Gestalt rules described above make it possible to develop Gestalt genetics. The latter studies holistic genetic patterns that are relatively independent of particular components. Gestalt genetics comes into contact with the teachings of the creator of analytical psychology C. Jung and his associate Nobel laureate in physics V. Pauli about the archetypes of the unconscious [17].

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

In addition to the phenomena of Gestalt psychology in living organisms, there are many genetically inherited physiological phenomena in which the same holistic picture is realized in conditions of a wide variety of constituent elements and which, by this property, can be attributed to Gestalt biology (this name combines genetically inherited Gestalt -similar phenomena of different types). For example, Gestalt biology includes some genetically inherited phenomena of morphogenesis (laws of phyllotaxis, spiralization of biological structures at different levels and branches of biological evolution), as well as some functional phenomena such as homeostasis at different stages of ontogenesis and the same type of processing of sensory information from different sensory organs in in accordance with the basic psychophysical law of Weber-Fechner.

Gestalt biology also includes the inherited phenomena of biomechanics of movements, described by the classic of biomechanics N.A. Bernstein: the general target task of movement is performed exactly regardless of the inaccuracies of its constituent motor subtasks [12, 18]. For example, when repeating an exact hit with a hammer on a nail, each time a person uses different trajectories, speeds and accelerations of body parts with changes in both the angles of flexion in the joints and the activity of many muscles of each joint with a wide set of motor neurons of each muscle (Fig. 2)

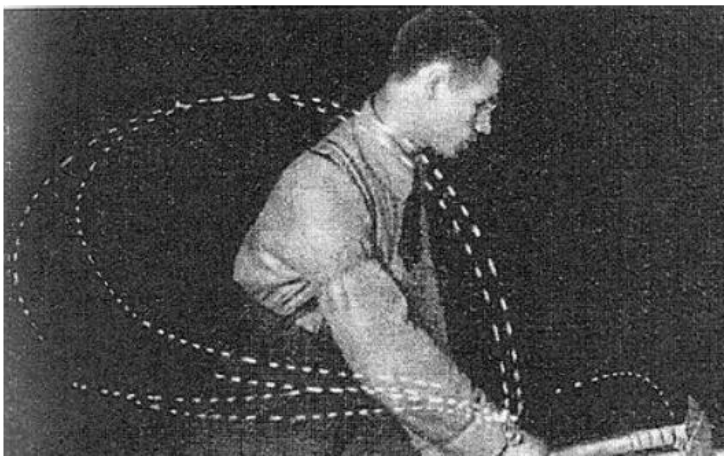

FIGURE. 2. Research of biomechanics of movements according to N.A. Bernstein (from the site https://ru.wikipedia.org/wiki/Бернштейн,_Николай_Александрович). Permitted to reproduce from Wikipedia under license: Creative Commons CC0 License

This question of how the central nervous system is able to adequately control the multiple degrees of freedom of the musculoskeletal system was first posed by Bernstein and is now known as the "Bernstein problem." The problem with degrees of freedom in motor control is that humans have different ways to perform a movement to achieve the same goal, using excess neurophysiological degrees of freedom. How the nervous system “chooses” a subset of these nearly infinite degrees of freedom is a major difficulty in understanding motor control and motor learning. In other words, under normal circumstances there is no simple one-to-one correspondence between a motor task and its motor solution. We hope that the further development of genetic biomechanics using the ideas of block-stochastic organization in living bodies will make it possible to effectively model and interpret such phenomena.

CONCLUSIONS

Representation of DNA nucleotide sequences in the form of a set of parallel n-texts, each of which is written in the corresponding DNA n-string alphabet, makes it possible to reveal previously unknown universal regularities of the block-stochastic organization of genomes of higher and lower organisms. In this case, it turns out to be useful to use the operations of the tensor product of vectors and matrices from the mathematical arsenal of quantum informatics and quantum mechanics, which brings genetic informatics closer to these sciences.

The discovered regularities of the multilevel block-stochastic organization of genomes consist in the relative independence of the percentage sums of n-plets that make up block tetragroupings in n-texts of genomic DNA from the values of the percents of individual n-plets in these tetragroupings. The corresponding genomic phenomena can be associated with the genetically inherited abilities of the brain, studied by Gestalt psychology, to form integral images that are relatively independent of their particular components. On the basis of these associations and analogies, it is possible to develop gestalt genetics as a science of holistic genetic patterns, relatively independent of private components. Gestalt genetics provides new opportunities for modeling and understanding genetically inherited phenomena of genstalt psychology and a number of other physiological phenomena and problems, including Bernstein's problem on the biomechanics of motor movements.

REFERENCES

1. I. Stewart, *Life's other secret: The new mathematics of the living world* (Penguin, New-York, 1999).
2. S V. Petoukhov, *Matrix Genetics, Algebras of the Genetic Code, Noise Immunity* (RHD ,Moscow,2008)(in Russian), available at http://petoukhov.com/matrix-genetics-petoukhov-2008.pdf.
3. S. V. Petoukhov, Biosystems **139**, 1-11 (2016).
4. S. V. Petoukhov, Biomashsystems **5(1)**, 78-145 (2021) (in Russian).
5. S. V. Petoukhov, Biosystems **209**,104503 (November 2021).
6. N. A. Bernstein, *The co-ordination and regulation of movements* (Pergamon Press., Oxford, 1967).
7. R. Alexander, *Biomechanic* (Mir, Moscow, 1970), (in Russian).
8. R. Glaser, *Essay on the foundations of biomechanics* (Mir, Moscow, 1988), (in Russian).
9. V. I. Filatov, *Clinical biomechanics* (Medicine, 1980), (in Russian).
10. V. I. Dubrovsky and V. N. Fedorova, *Biomechanics* (VLADOS-PRESS, Moscow, 2003), (in Russian).
11. G. Brankov, *Fundamentals of biomechanics* (Mir, Moscow, 1981), (in Russian).
12. N. A. Bernstein, *Essays on the physiology of movements and the physiology of activity* (Medicine, Moscow, 1966), (in Russian).
13. M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, New York, 2010), DOI: 10.1017/CBO9780511976667.
14. F.S. Perls, *Gestalt approach*. (Psychotherapy, Moscow, 2007), ISBN: 978-5-903182-19-0.
15. N. Lebedeva and E. Ivanova, *Travel to Gestalt. Theory and practice. Textbook* (Speech, Moscow, 2015), ISBN: 978-5-9268-1877-9.
16. J.-M. Robine, *Gestalt therapy* (Institute of General Humanitarian Research, Moscow, 2007), (in Russian).
17. G. Wehr, *Carl Gustav Jung*. (Ural Ltd, 1998), (in Russian), ISBN: 5-88294-073-7.
18. J. M. Feigenberg, *Nikolai Bernstein : from reflex to the model of the future*, edited by Eberhard Loosch and Vera Talis (LIT, Berlin, 2014).