Fractal Structure in DNA Code and Human Language: Towards a Semiotics of Biogenetic Information

Peter P. Gariaev*, Boris I. Birshtein*, Alexander M. Iarochenko*, George G. Tertishny*, Katherine A. Leonova*, Uwe Kaempf **, Peter J. Marcer***.

*Wave Genetics Inc. 87 Scollard Street, Toronto, Ontario, Canada, M5R 1G4,
gariaev@wavegenetics.com;
** Institute f. Klinische, Diagnostische und Differentielle Psychologie-
AmFalkenbrunnen - D-01062 Dresden TU, Dresden, Germany,
uwe@psy1.psych.tu-dresden.de;
***55 rue Jean Jaures, 83600 Frejus, Var, France,
aikidopeter@aol.com, http://www.bcs.org.uk/cybergroup.htm

Abstract
Evidence obtained by computer analysis confirms the scientific validity of the now wide-spread use of semiotic-linguistic nomenclature to describe the workings of the genetic apparatus. It shows that both natural languages with their semantic speech structures, which are sign representations of human mental activity and thought, and genetic texts(ie DNA/ RNA and proteins, etc) exhibit a strategically close geometrical fractal framework. A similar analysis of random texts of the same characters leads to the loss of this frame-work. This not only points to a causal supergenetic relationship between such texts that proceeds at some level in the demonstrated fractal framework, but that, for example, Chomsky's concept of a universal grammar, in relation to all forms of human languages (including mathematics and computer languages), is probably not only correct but is naturally genetically inherent.

Keywords: semiotics, linguistics, universal grammar, DNA-wave biocomputation, genetics.

1 Introduction

Detailed evidence from a linguistic perspective is presented in further support of conclusions reached in earlier paper entitled the DNA-wave biocomputer [42]. There, experimental evidence, computer analysis and theory showed that DNA action is that of a "gene-sign" laser and its solitonic electro-acoustic fields, such that the gene-biocomputer "reads and under-stands" the DNA texts in a manner similar to human thinking, but at its own genomic level of "reasoning". It asserted that genetic texts and natural language texts (irrespective of the language used) have similar mathematical-linguistic and entropic-statistical characteristics, where these concern the fractality of the distribution of the character frequency density in the texts. The new analysis presented below therefore fits well together with this previous work (presented at CASYS and published in the Proceedings) supporting this hypothesis of the DNA wave biocomputer, that DNA, conceived here as the semiotic gene-sign continuum of any

International Journal of Computing Anticipatory Systems, Volume 13, 2002
Edited by D. M. Dubois, CHAOS, Liège, Belgium, ISSN 1373-5411 ISBN 2-9600262-7-6

biosystem, is able to form quantum holographic/solitonic pre-images of its organism, and that this continuum is the measuring, calibrating field for the organism's construction. It also says, as has been shown elsewhere in relation to quantum holography [43], see section 3, that not only must there exist, as is the requirement for all computation, a canonical labeling of these pre-images in terms of a semiotic sign structure of the genome (ie a universal grammar, Chomsky's hypothesis!), but that the fractal of the sign structures of the genome must reflect its and its organism's quantum nonlocality. A nonlocality at six levels of genetic/metabolic information – tissue, cell, nuclear, molecular, holographic and quantum, which, it is again argued here, provides the supercoherence for multicellular organisms and their unlimited evolution.

2 Some Linguistic Concepts Applied to Genetic Texts

The use of a semiotic-linguistic nomenclature in relation to the working of the genetic apparatus, that has already become traditional, is as yet of a mainly metaphorical character. In scientific, and even more so popular literature such fundamental linguistic concepts as "texts", "reading" (for example, a ribosome "reads" an mRNA), "compilation" (translation of mRNA-sequences into a protein), "punctuation" (codons with a starting and a stopping functions), "word" (nucleotide sequence, coding proteins) etc, have become commonplace, and such phrases as "the language of the DNA", "the language of the protein" are widely used. However while such vocabulary is quite convenient and useful, it creates the illusion, that in this area of genetics and biology everything is well understood.

This raises the question "Can this use semiotic-linguistic nomenclature be scientifically justified?" For taken at face value, from the semiotic logic of these terms there emerge more binding ones, for example, "silent genes", "a recognition of the promoter" [1] etc. It then becomes logical to ask, who or what is/are generating such nucleic "texts" and performing

their "reading" in a living cell? This linguistically specialised nomenclature has thus become a corollary of the silently accepted notion of a genome, which is thought to operate as a quasi-thinking cellular structure, some kind of a biocomputer. In fact such an approach is probably justified and it is the one taken in the present research as well as in recent papers [2,7,42]. Other authors also take same position, for example, within the framework of hypotheses regarding a molecular-genetic control system [21].

However such an origin of actual genetic texts or sequence of semiotic signs in genetic molecules, as for example, the programs of a biocomputer renders references to the Darwin's evolution theory as insufficient. A semiotic approach to DNA as a biocomputer which "learns", as offered by us and other authors [2,7,21,42], is probably more fruitful. The present research is thus a new attempt, following our last paper [15], to define, whether such terms of modern linguistics and semiotics (i.e. "sign", "word", "morpheme" etc.) should indeed be applied to some sequences of DNA and RNA. Its theoretical basis is an ecological-semiotic model of language, one of situated adaptation and self-organization [23-25], where systems of a special sort of natural classification exist that could be applied to genetic languages as well. This supposes that in living nature these classification processes reflect the environment firstly in form of images, then in form of their comparison, sequential summations, and at last generalization (i.e. the selection and the emphasis on common and recurrent elements as well as the omitting of specific parts within the contents of the initial images). On this basis, images are built, which are more abstract, and, in their turn, firstly form a particular hierarchy, a sort of "pyramid" of images with an extending size of area that relates to each of the more abstract images. Secondly, certain aspectual subsystems of abstraction are derived. These processes form the basis for orientation, for the survival of any living being in an environment, appropriate to them. The typical set of situations of vital activity regarding the individuals of the given species then compels them to form in their reflecting sphere a certain system of hierarchically related abstract images. The most specific and the most concrete images correlate with the concrete sorts of situations of vital activity, whilst the more abstract images reflect the most common, steady and recurrent features of situations which relate to the types of situations.

Thus the situations of practical vital activity shape the systems of practical natural classification, and the situations of communicative vital activity shape those of communicative natural classification. The communicative situations are opposed to practical situations, since the former formulate and achieve objectives which are not of an immediate practical significance (providing food, finding shelter, flight from predators etc.), but which comprise a "transmission" of the information to communication partners, for example, during the postulated intercellular wave communication.

Classifications, which originate under such conditions, consist of special communicative abstractions called values. The function of this type of classification means that, through the activation of units belonging to a classification, within the reflective system of the partner (achievable through sending some intermediaries, i.e. physical objects called signs), the partner is enabled to send signals about any specific, practically significant meaning. Thus a sufficiently high level of an abstractness of values with their relative scarcity in comparison with their meanings, helping maintain a relative identity for each member of a social system (including a multi-cellular continuum of biosystems), enhances the unity of the language or of a totality of languages achievable in a community on the basis of mutual dialogue and the training of its members (be it people in a society or cells in a biosystem).

The difficulties caused by the application of the general semiotic terms, which have been worked out mainly through the analysis of social communicative systems, to genetic systems of communication, and, in a more general case, to a metabolic system of communication already arise at the stage of an identification of senders and receivers within them, and, all the more, concern recognition of those different signs and their functions, which are utilized in this process. On the one hand, a sender, i.e. a subject of the genetic communication, can be considered a cell genome, which - under different condition - in different environments or at various stages of cell development (and, probably, within the multi-cellular organism, where the cell is included) - maintains its regulative and communicative activity (to be precise, a regulative activity by means of a communicative one) towards other cell subsystems. In this case the latter can be considered as receivers (remember the analysis of wave sign images at a level of DNA, collagen and ribosome, given in the first paragraphs).

Of course, in such cases the most difficult problem is a segmentation of genetic structures into speech-alike (sign) sequences, the detection of their functionally relevant combinations of sign units of a different level (similar to the detection of morphemes, words, word collocations, sentences etc. of a natural language) as well as a definition of their standard functions (values), and of the limits of the "contextual" implementation of these standard functions and, eventually, of the relevant context types.

One of the seeming evidences that DNA-"texts" are texts in a semiotic sense (a main feature of which are sequences of words, united through meanings) is the correspondence between the allocation statistics of the DNA-words and the law of Zipf, Katsikas and Nikolis, as, for example, detected in a weak correspondence between the allocation of purins and pyrimidines in DNA-sequences [6]. However, as the paper of B. A. Trubnikov [22] shows, this law refers not only to biological objects. Among those objects, which are well described by it, are also such ones as distribution of stars according to their masses within star constellations, allocation of chemical units according to their masses in an earth cortex etc. Therefore the compliance with the Zipf's law cannot be the only criterion for the selection of word-like units from the DNA-texts.

But at the same time this law points, by means of selection in accordance with this feature, to a definite fraction of sequences of DNA, which take the part of "words" [15].

Further there is seemingly evidence in favor of the typological relationship between words of human speech and DNA-"words", demonstrated in some studies on "linguistic genetics", investigating overall laws of language development [8]. Thus the word formation within the onto- and phylogenies of language appears to comply with the laws similar to those of formal genetics, including splitting of morphological and semantic features, transmitted from "parent words" to "hybrid words", with dominant and recessive features, with transpositions of word radical structures by mutations etc. However, this cannot be considered as a sufficient argument for the semiotic nature of DNA, since the quantitative laws of formal genetics can be explained on the basis of the statistical-combinatorial mechanisms which are not a specific property of biosystems only.

The study of the genetic communications as an anagogic counterpart of the speech process requires - apart from findings within the such sequences of DNA-nucleotides as macro-features, which are similar to those of the human speech - at least a fragmentary confirmation of the operation of sign and allusion mechanisms in the genetic communications. This necessitates segmentation, i.e. a search for natural boundaries between various sites of genetic messages (signs of a different sort and level), to find those functions (values), which presumably belong to each of these signs etc.

Also important is the notion that DNA fulfills administrative functions at a level of the control over its organism's development, where in human speech, coding ideas, doctrines and laws, fulfill the same control functions, which are fractal and so enlarge up to the level of a public organism, which grows, develops, merges with other public organisms, creates descendants, grows old and dies along with its components, i.e. people. Such social-genetic processes, which are implemented through one of the main sign systems, ie through speech, require further studies.

The idea of a quasi-verbal or, what means the same, of a figurative level of DNA code functions (within the limits of biosystem's chromosome continuum) opens the way out of a restricted sign continuum of the triplet genetic code which is unable to explain how the main contents - the space-time structure of the organism - are encoded in the genome.

The ultimate goal of the offered analysis is the comprehension of "word" and "sentence" semantics in structurally functional space and sign space between DNA, RNA and protein, which - in case of enzymes at least - is extremely heterogeneous (fissile center, sites of recognition, architecture of hydrogen-hydrophobic forces of the self-organizing of a peptide chain). The multilingual metabolic "talk" between cellular information biopolymers and their functions as the exchange outcome between quasi-verbal biosignals suggests two mutually correlating levels of this exchange, namely a material and a wave level. The material level is well-known (matrix copying DNA, RNA and protein, interaction antigen-antibodies, self-assembly of cellular structures), whilst the wave level, which is closely related to it, has been investigated very poorly. In the latter case the situation is not simple, but no less significant. The electromagnetic and acoustic radiation of proteins, nucleic acids, and cytoskeleton are well-known. It is thought, that it is a wave level of information contacts of cellular-tissue space, which gives the metabolic processes a field dimension with its language specification and regulation [2,3,7,9].

The considered bioinformation currents, being linked to a metabolism and energy, are not restricted to division of sign sequences into a substance and a field, but rather they are extended by means of a fractality of these sequences. For within the acoustic-electromagnetic component of DNA signal functions, the fractality of a soliton field is observed, which maybe formally described by equations within the framework of the phenomenon of Fermi - Pasta - Ulam[2,7,9,27-29,42]. It makes the semantic analysis of protein- nucleic and other information contacts of biostructures even more difficult. We assume therefore, that in living cells there is a hierarchy of substance-wave sign structures, where the relative gradation "character (phoneme) - morpheme - word - sentence... " predetermines the fractality of these structures. That is to say, a "word" on one scale can be only a "character" on another, as is often the case in mathematical and computer languages.

Other difficulties are related to the concept of a "framework of reading", for example, by decoding of mRNA codons through a ribosome. A shift by only one character in a triplet of mRNA (or, at a wave level, by a small change of a phase, polarization, frequency of electromagnetic and/or of acoustic radiation of genetic structures) could completely interchange the meaning of the text being read (ie a received electromagnetic and/or acoustic image). For example, texts of the same sequences of DNA could correspond to different languages. Nevertheless, some definite logic of handling metabolic (and wave) bioinformation is inevitable, if we are to understand the essence of a phenomenon of Life. The above mentioned concept does not refers only to the known canonical triplet genetic code. The code is however a convenient starting point, when the initial level of such proposed DNA-polycodes - the level of substance-matrix gene signals, which makes up 1-5 % from all mass of the genome DNA, is deciphered. But such polycodes mean that the remaining majority of DNA, often considered by geneticists as "garbage", probably carries strategic information on the biosystem in form of wave biosigns of the soliton, and other figurative sign structure, that might include a speech-similar structure too [2,5,7-11,12-15].

It is possible then to address this problem from the standpoint of a fractal representation of genetic "texts", in a so-called Chaos Game Representation (CGR) of languages. In accordance with the CGR concepts, the method of compact and visual fractal-graphic representation of nucleotide sequences of DNA was offered [18]. The procedure of such a mapping of DNA-"texts" is described as follows. The nitrogen bases, adenine, guanine, cytosine, thymine of DNA are represented by each of four quadrants, as designated by its center point (ie the "characters" A, T, G, C) see Fig. 1, such that the first base in any sequence of DNA bases is represented by a point set on a middle of the segment that links the quadrate center of the four quadrants to the center of the quadrant of the first base. Each subsequent base is then represented by a point, set again in the middle of the segment that links the previous point with that of the center of the quadrant of that base (see. Fig. 1).

Thus some essential properties of a CGR-representation are :-

Property 1. Each sequence has only one CGR-representation; different sequences have different representations.

Property 2. For any generated point of the quadrate, it is possible to specify a measure for that sequence to that point, by whatever is the small distance reached beforehand.

Property 3. By reviewing summations of the representations of each possible sequence from the alphabet (A, T, G, C) this totality belongs to the self-similar set with dimension of a similarity $d = 2$ (see [17]): $d = - \ln N / \ln r(N) = - \ln 4 / \ln (1/2) = 2$, where N is the number of diminished copies of quadrate, with which it is covered, r(N) - coefficient of scaling smaller than 1, i.e. this set " fills in the whole quadrate " (is like the Peanuts curve which also has fractal dimension equal to 2).

/Remark:/ This algorithm can be considered as a proof among the others that the number of points contained in a segment is "equivalent" to the number of points contained in a quadrate.

The main idea of the CGR may be used for the graphic representation of languages whose alphabet has more then four characters. Into consideration can be taken the CGR-representation of any natural language or CGR-representation of the language of amino-acidic sequences of proteins [15]. I.e. in this case, a more advanced variant of the CGR-representation for languages with any number of characters in the alphabet is suggested which complies with the properties 1 - 3. This approach generalizes the notion, used in the present paper. Let us consider first the case of nucleotide alphabet comprising four characters. We divide the quadrate into four subquadrates (in this case there are four subquadrates or quadrants) and then set the alphabet characters A, T, G, C. in relation to them.

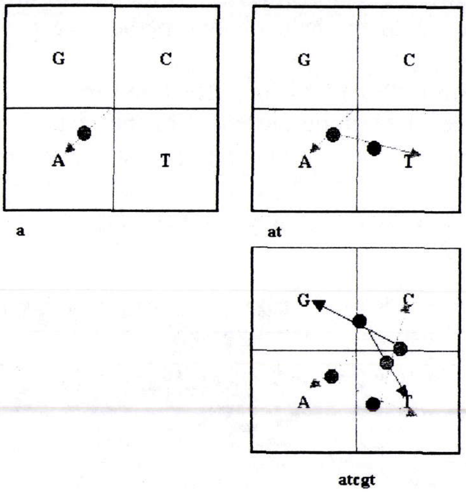

Fig. 1. Scheme of the CGR-representation of the line "atcgt" in the nucleotide alphabet (A, T, G, C).

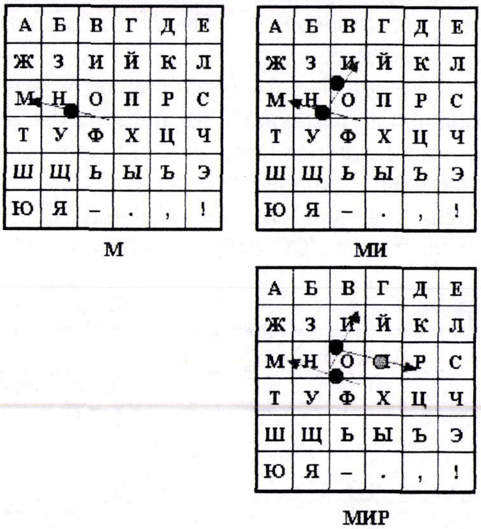

Fig. 2. Scheme of the CGR-representation of the line "World" (in Russian «МИР»), Russian alphabet.

Each of the quarters/quadrants is similar to the total quadrate, therefore it can be represented in any of its quarters by means of parallel transfer and scaling at a coefficient r = 1/2. This representation is mutually non-ambiguous. The "empty" chain of characters will correspond to the center of quadrate. Each subsequent character determines the representation of the whole quadrate in an appropriate quarter. A graphic representation of this character will be an image of a graphic representation of the previous character; in particular, the representation of the first character will be the image of the center (see Fig. 3)

In case of the four character alphabet this algorithm evidently leads to the same outcome as the algorithm used in [18].

If we change the coefficient of scaling (for example, r =1/6) and, accordingly, the number of diminished quadrate copies, which cover it (in this case there are 36 copies), we can receive a graphic representation of the text, for example, in Russian, by putting in correspondence to each of 33 characters of the Russian alphabet one of the subquadrates. The next character determines as before, a mapping in the appropriate quadrate; a graphic representation of this character will be, as before, an image of a graphic representation of the previous character (see fig. 2).

In biological studies [19,20] the suggestion was made to apply the CGR-representation for searching function sites of the DNA. According to each known set of genes a kind of recognition matrix is built, which is called by the authors "a mask of the fractal representation of a set" (mask FPU). This mask defines a measure of the closeness between the given nucleotide sequence and the set of sequences. And in comparison with other known methods this mask method is simple to implement and has the advantage of efficiency (the amount of operations depends linearly on the length of a recognized sequence). The outcomes obtained with the help of this method are now demonstrated.

It is argued that the mask method is useful in both natural linguistics and in genetic linguistics, particularly for defining the degree of semiotic closeness of the texts. For example, we took three texts of nearly identical size (about 300 thousand appearances of characters from the Russian alphabet).

Figures 3-5 show that the CGR-representations of the texts 1 and 2 are similar, but both differ from the CGR-representation of the text 3.

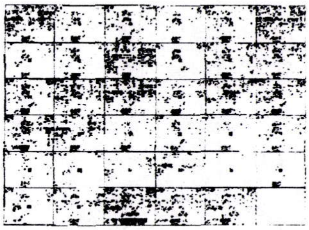

Fig. 3

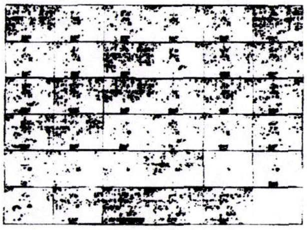

Fig. 4

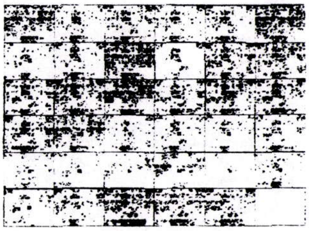

Fig. 5

Fig. 3. CGR-representation of the Text 1 (Stephens A. A technique of programming by Turbo C.), Russian alphabet. Global fractal dimension D1 = 0.71,

Fig. 4. CGR-representation of the Text 2 (Shildt). C for professionals.), Russian alphabet. Global fractal dimension D2 = 0.73,

Fig. 5. CGR-representation of the Text 3 (Gariaev P. ,1994, Wave genome.), Russian alphabet. Global fractal dimension D2 = 0.87,

where to define the fractal dimension, the algorithm of cell count (see e.g. [17]) was used, and to simplify the process, the CGR-representation on an one-dimensional, instead of two-dimensional carrier (i.e. on a segment instead of on a quadrate).

The fractal dimension of their CGR-representation is thus an essential characteristic of the texts. In our example the situation is as follows: the global fractal dimensions of the CGR-representation of the texts 1 and 2 are close (D1=0.71 and D2=0.73 accordingly), but they strongly differ from the global fractal dimension of the Text 3 (D3=0.87). Furthermore the fractal dimension of CGR-representation of a "random" text (i.e. sequence of characters, where each subsequent character is chosen with the same probability and irrespective of previous ones) is equal to 0, and the fractal dimension of a CGR-representation of texts consisting of a sequence of repetitions of one arbitrary character, for example, "aaaaa..." is equal to 1. Therefore the fractal dimension D of the texts written in natural languages varying in a range $0 < D < 1$ can be considered as a certain measurement of the complexity of a semiotic structure of the compared texts (and perhaps of the complexity of their meaning). In this case too, the structure of the compared texts is only at a level of their character sequences, which, in turn, by virtue of the fractal ratios in natural language texts, is derived from the structure of verbal sequences, and, eventually, from relative richness of lexical vocabulary of the texts and complexity of its contents. This opinion is confirmed by our evaluation of the word form and lexeme dictionary in case of each text. Thus our postulate of the possibility of an integral evaluation of the semiotic object according to its characteristics received on the basis of the analysis of elementary units of its structure (in this case, characters) is confirmed.

This conclusion has also a major methodological validity for the comparative analysis of such super-composite sign objects as DNA separate texts and genomes of different taxonomic groups as a whole. Presently, the genetic structures as semiotic units are mainly investigated at a level of triplet genetic code. The majority of its remaining units and organizational levels remain largely uninvestigated. Thus as yet biology has a lot to do until the picture of a semiotic structure of the genetic apparatus can be considered as relatively complete and understandable. It is therefore difficult from the perspective of semiotic analysis to understand the excessive optimism shown, for example, in the paper [21] which deals only with "words" of genetic code triplets. However, the methodology offered here deals not only with such "word triplets", it enables the comparison of different sequences of DNA and RNA providing an evaluation of the degree of their similarity and difference as well as a measure of the relative complexity of their sign structure [26].

To illustrate our method of fractal representation of natural and genetic texts the density matrixes for a text in English (manual on computer programming) (Fig. 7) and a "text" of a casein gene (milk protein) (Fig. 6) are shown.

Fig. 6 Density matrix of a random-game representation of a nucleotide sequence (gene) in the alphabet (A, T, G, C), coding an initial structure of the casein (milk protein).

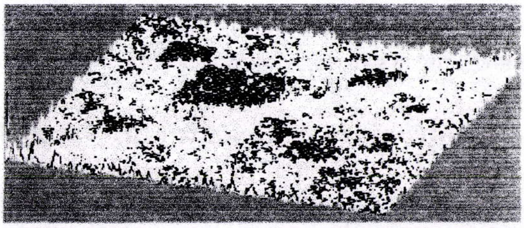

Fig. 7 Density Matrix of a random-game performance of the text in English (manual on computer programming). The structure of appearances in the text of four parts of speech was considered. The left front corner corresponds to the word "the", the right front corner to the word "in", the left far corner to the "on" and the right far corner to the "of".

This method provides an essentially other possibility of quantitative and qualitative matching of natural and genetic texts. A similar outcome can be obtained in another way, and also through a new method, as it is shown on the graphical histograms of a similarity and differences among flanks and introns of a major gene group (Fig. 8). By the same method the histogram of matching of natural texts (Fig. 9) was obtained regarding the monograph of the author "Wave genome" and stories of the Russian writer F. Abramov [Unpublished findings of the co-operative research with the co-author M.Y. Maslov (Mathematical institute of the Russian Academy of Sciences/RAS).

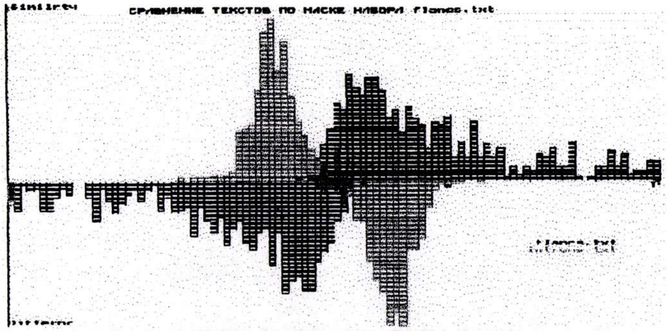

Fig. 8 Comparison between flanks and intrones

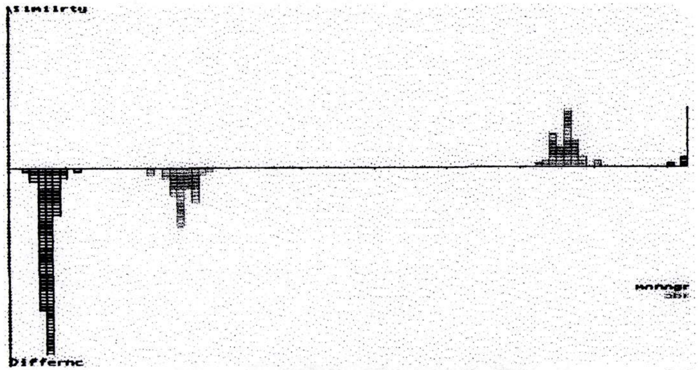

Fig. 9 Comparison between the monographs: Gariaev P. Wave genome, Moscow, 1994. and Abramov F. True and untrue stories, Moscow, 1993.

3 Unity of the Fractal Structure of Genetic and Natural Texts

Here a closer look is taken at the fractal semantic niche of sign processes in the genetic apparatus of higher biosystems, linked to its quasi-speech characteristics and to the genetic attribute system of word forms in natural human languages. Earlier, some proof was obtained, that the development of the languages and human speech submits to the laws of formal genetics [30]. It showed that DNA "texts" (quasi-speech) and the written culture of the humanity, as in real speech, essentially fulfill identical administrative, regulative functions, but at different fractal scaling. But whilst DNA functions genetically at a cellular-tissue level, human speech being a macro genetic structure is used at a level of the social super organism. Therefore the prior metaphorical usage of linguistics concepts in relation to DNA, when the terms "word", "text", "punctuation", "grammar" were used arbitrarily and attempts were made to understand other semiotic dimensions of the genome intuitively, has now been , in a certain sense at least, partially validated. It is supported by application of the fractal theory and the method of code conversions to DNA sequences and the human texts structure, leading to the conclusion that the DNA and human speech (texts) have a strategically close fractal structure in relation to geometrical, and, possibly, semantic expression. It, too, seems to correlate with the fractal frame of the soliton acoustic and electromagnetic FPU-field, generated by the chromosome apparatus of higher biosystems. A concept proposed in our earlier experiments with plant genomes [27-29].

The outcome of these experiments was interpreted by us from the viewpoint of the semiotic-wave component of the genetic code, and has an essential methodological value both for the analysis of such super sign objects as the DNA texts as well as for the genome as a whole. This could therefore open up totally new areas of the chromosome apparatus meaning.. However, biology and the gene-linguistics have a long way to go yet to get a clear and semiotically understandable picture of the DNA sign sequences. Our way of interpretation of genome functions allows for the comparison of different natural DNA and RNA sequences with the evaluation of the measure of their similarity and difference as well as with the degree of relative complexity of their sign structure. But even more important, a new method of the comparison of semantic constructions of human speech and the coding DNA sequences emerges. If it is right, new dimensions of comprehension of thinking and consciousness can be expected by their reflection in sign (semantic) sequences at different levels of living matter organization, namely at a level of human speech (a higher form of consciousness) and at a level of quasi-speech of genetic molecules (ie a possible quasi-consciousness of the genome). This evidence fits well with the mathematical-linguistic model of Chomsky, that postulates common principles, which underlie any language and which join into one "universal grammar" [31]. From Chomsky's view, such "universal grammar" is inherent, i.e. it has some genetic determinants. This is an extremely important circumstance, which once again emphasizes the supergenetic relationship of the DNA semiotic structures and human speech structures. To a limited extent we have confirmed this position in the present study (see earlier mentioned paragraphs) by showing the similarity between the DNA and the human speech. Chomsky is probably right, arguing, that the in-depth syntax constructions which make out the basis of the language, are passed down from generation to generation, providing each individual with the capacity to learn the language of its ancestors. The fact that a child easily learns any language is then explained through the theory that the grammars of all languages coincide. The essence of the human language is invariant for all people. But it can now therefore be supposed, that this invariance is spread even more deeply, down to the macromolecular semantic ("speech") chromosome structures. Further independent confirmation in relation to the DNA-wave biocomputer comes from quantum holographic imagery [42,43]. For here, 3 dimensional spatial object images, the observations, are phase conjugate (pc) so as to coincide with the 3 dimensional objects themselves, the observed. That is to say, it is the 3 dimensional objects themselves that implicitly label all aspects of experience, the observations, in a universal way for all observers, so as to form the basis of communication between all those observers with a common genetic heritage and sensory apparatus. The bases of all languages in this case are therefore shared arbitrary semiotic labelings of these object and their properties such that

Each (pc) holographic encoding on a 3D object or icon	maps to and from < --------------->	the holographic encoding of the arbitrary label for that object or icon
right	corpus callosum	left

where such mappings are unique since no two objects can occupy the same position in 3 dimensional space. This mapping schema could then explain the morphology of the human brain which concerns the two brain hemispheres and the corpus callosum, which joins them. That is to say, the right hemisphere that realizes the holographic encodings of the real world, is the artistic brain, and the left, that realizes the arbitrary labelings of these real world objects and their properties, is the logical brain. For in the latter, an essential element of the mapping of such labeling of objects includes numbers and sets and their logical relationship one to the other, where these must be acquired by learning. This mapping schema can therefore be postulated as the basis of Chomsky universal grammar.

Such a semiotic approach to genetics is further confirmed by certain theoretical-experimental evidence we obtained [27-29]. This evidence points to some key methodological approaches so as to reveal as yet unknown semiotic layers of the genetic apparatus of higher biosystems, which includes, it must be noted the sign-field environment of the Earth's Biosphere. This is a strategic environment where small semiotic-wave distortions could have profound effects indicating the need for great care, whilst the ideas of the wave (and "speech") genetics are still in a early phase of their active investigation and development. A system of strictly determined prohibitions on particular experiments in this new area of knowledge, such as already exists in gene engineering, for example, on cloning of higher organisms, is therefore almost certainly necessary.

The ultimate goal of the offered analysis is the selection of sign units of different levels and the comprehension of their semantics in the function space „DNA - RNA - PROTEINS- ORGANISM“, which, at least for enzymes, is extremely heterogeneous (fissile center, sites of recognition, architecture of hydrogen-hydrophobic forces of the self-organization of the peptide chain). The multilingual metabolic "talk" between information biopolymers of a cell and their functioning as an outcome from the exchange of sign biosignals suggests two mutually correlated levels of this exchange - a material and a wave level. The material level is well investigated (matrix copying DNA - RNA - PROTEINS, interaction antigen-antibody, self-assembly of cellular structure), but, as already mentioned, the wave level closely related to it, has been not investigated by the establishment science. The situation here is difficult. The electromagnetic and acoustic radiation of protein, nucleic acids, diaphragms and cytoskeleton needs intense and careful study. For if as shown here, they are sign radiation, then this level of information contacts of cellular tissue space transfers the metabolic processes into a physical wave semiotic dimension having its own "language" specificity and regulation. Further such bioinformation currents, linked to a metabolism and energy, will not be restricted to dividing of the sign sequences into substance and field, but will also extended, as has been discovered above by means of the fractality of these sequences. For example, in the acoustic-electromagnetic component of the DNA signal functions a fractal of the soliton field is observed, which can be described formally by the theory of the phenomenon of Fermi-Pasta-Ulam (FPU). It even more complicates the semantic analysis of protein↔nucleic and other information contacts of biostructures. It is thus possible to postulate with some certainty that in living cells there is a hierarchy of substance-wave sign structure, where the conditional gradation "character (phoneme) - morpheme - word - sentence... " causes the fractality of these frames. It is urgent therefore in our view to begin to comprehend these mesomorphic "texts" of chromosome DNA, readable in three- or n-dimensional space, when "characters of words" and "words and the sentences" themselves are drawn up not only in one line and in one dimension, but are being drawn up and "read" backwards and forwards, upwards and downwards, etc. In such dynamic processes (with returns to the same "texts" by means of an appearance of FPU [27,28]) the infinite continuum of anisotropic "filaments of the texts" is built and deleted, which spreads in all directions of the mobile interphase chromosome continuum of the whole genome space of the biosystem.

A connected phenomenon which seems to be related to all above discussed features of the genome of higher biosystems is the origin of life, particularly on Earth. *It has been discussed for a long time. There are a lot of theories.* Consider the hypothesis of panspermia, but not in the variant, which argues that certain spores - ancestors of all biotic forms - were brought to Earth. But rather the process of the natural evolution of abiotic organic molecules, predecessors of DNA, RNA, protein and other essential parts of biosystems, originated from "an initial broth" combined with an act of introduction of exobiological information into the first nucleic acids. This information could be speech-like, and its structural essence fractal. It began by an introduction of a doublet - triplet DNA-RNA code, as the first stages is an elementary language with four letter alphabet. Then it transformed into a 20 character protein alphabet and into further higher languages in the sense already discussed. Generally, the hypothesis of a prehistory of the initial DNA language is widely discussed, beginning with the breakthrough study of V.I. Scherbak [32]. He proved that the probability of an evolutionary provenance for symmetries of the genetic code is equal to zero. The beauty and elegance of his theoretical analysis where for example such parameters as the nucleon relations in amino acids and degeneracy of the genetic code are used, supports our reasoning. However there is one qualification, the need to take account of the fractality discussed here. Since the actual genetic code has the leading wave sign component and cannot simply be reduced to a programming of protein synthesis [27-29], the concept of a prehistory, too, of the code also requires a certain specification. In a real (fractal-heteromultiplet) code the existence of other artificial sign mathematician frames - compared with ones discovered by Scherbak, can be foreseen.

Developing this notion and ideas, as suggested earlier, it is argued that would be a naive simplification to consider DNA "languages" and "writing"("speech") to be a complete analogy of human verbal structures. It would be more precise to suggest, that the DNA functions are based firstly on its meta-language with its own grammar genome. In this case the appropriate analysis of meta-languages provided by À. Solomonik is extremely useful [33] and can be applied to mathematics. This analysis leads to the conclusion, that mathematics is a meta-language too, having a code of rules of structuring its texts. In contrast with the conventional speech, where a phrase, expressing a particular thought, can be built in dozens different ways, verbal (sign, fractal) sequences in mathematics are generated by means of a small number of strict rules. In an automatic mode the latter allow the reception of a prognosticating outcome, that "predicts" for example a type of resonant interactions between the external sign exobiological physical fields with the information bio-macromolecules of DNA, RNA and proteins [29].

This example shows, how on the final stage of mathematical, meta-lingual as well as grammatically strongly oriented conversions of the texts, the result is gained in form of a physical-mathematical image of potential behavior of biosystem's information elements (DNA, RNA and proteins) under influence of external sign electromagnetic and sound radiation. The chromosomes too, probably, operate with meta-languages to create an "ideal" (physical-chemical-mathematical) model of a biosystem, an inaccessible prototype of a real organism. Such model can be called a wave supermatrix for an growing organism. Also, such model is to be more informative in comparison with, for example, a purely holographic model of a biomorphogenesis [27], and to supplement the latter.

The DNA, the chromosomes of organisms on the Earth are simultaneously donors and acceptors of not only its own wave commands, but of certain external (perhaps, exobiological) semiotic wave influences too. Partially we have shown this earlier. If we take it into account, we need to realize, that a new artificial, made by the people, electromagnetic semiotic entrance channel into the noosphere and the gene pool of the planet Earth requires the highest attention as to the good sense and expediency of our supergenetical manipulations. For these are - in fact – uncontrollable, since in this case they enter into competition with probable exobiological wave control over the gene pool of the planet. Perhaps, we have even entered a bifurcation point to choose a strategy of mankind's evolution: to go further and perhaps even recklessly on such a technogenic path, or to learn from our own bodies, into which the semiotic knowledge of the nature is embedded? Is it useful and necessary? There surely can be no clear answer to such questions at present, but is now imperative to ask them.

4 The Emergence of a New Semiotic

The evidence presented of the emergence of a new semiotic, from perspective of fractality and the position of quantum nonlocality of the genome : the Einstein-

Podolsky-Rosen effect in the activity of the genetic apparatus, requires a possible modification of the central dogma of molecular biology.

In 1935 A. Einstein and his colleagues B. Podolsky and H. Rosen [34] initiated an investigation - a thought experiment into the quantum relationship of elementary particles - the result of which is reducible now to the following. A quantum object, which is the fission of , for example, two so called "entangled" photons, maintains a new kind of informational link, such that the quantum status of one photon, e.g. the state of polarization (or of the spin in the case of two electrons) can, under the right circumstances, be instantaneously, i.e., in zero time, transmitted to the other photon. During this event, not only can the photons be at any distance from each other, but the state of first photon becomes unknown. Further since this distance can be greater than that required by the speed of light for a classical transmission of information between the two photons, this phenomenon is called the paradox or channel of Einstein-Podolsky-Rosen (EPR) often referred to as "quantum nonlocality" so as to underline the instantaneous distribution in space-time of the states linked by means of the quantum-nonlocal channel. It seemingly implied too in 1935 that the fundamental principle of causality, that signals all must travel at the speed of light or lower earlier established by Einstein would be broken, but many years were to elapse before this dichotomy between the two theories could be resolved. For firstly it needed D. Bell's updated study [35] of the EPR phenomenon so that an actual quantum mechanical validation or quantum measurement of it could even become experimentally possible, and the quantum simultaneity of the "entangled" status of fundamental particles experimentally proven. More years were to elapse, however before Bennett and his colleagues argued, that such particles can serve in their entangled space as mutual "information carriers" of their states, in relation to the phenomenon now known as quantum teleportation. Its experimental proof was realized by research studies of two groups – an Austrian team headed by Anton Zeilinger [37], and independently by the Italian team of Francesco de Martini. These discoveries and their experimental validation have been fundamental to research programs to build quantum computers based on these new informational principles, where their operational speed, can be shown to entirely exceed those at all possible on existing computers.

It can therefore in relation to chemical and thus quantum mechanical systems such as the semiotic sign systems and texts of DNA molecules and indeed human brains, be postulated, that these biosystems will necessarily incorporate such novel forms of information processes and transmissions since these new quantum modes clearly offer substantial evolutionary advantage. Furthermore, the fractal properties of the genome show that the properties of these semiotic texts, which belong to the microlevel of its organism will be reflected on many scales up to the level of the whole organism including of course the macroscopic. A property which in relation to quantum nonlocality implies the instantaneous governance of the workings of the organism at many levels including the whole. Indeed the semantic semiotic structures of the genome of multicellular biosystems would have the ability to propagate instantaneously through the chromosome continuum of biosystems [36].

It is postulated, on the basis of this fractality that the phenomenon of quantum nonlocality is used by biosystems on a chromosome level as one of the key factors of self organisation. This is rather attractive both in a philosophical as well as in the pragmatic sense. Such ideas corresponds well to the data about the wave/sign (semiotic) assignment of gene-information-metabolic and even the mental areas of biosystems. In our field versions of genome activity [27, 29] the EPR-effect is valuable one, which can explain the chain of semiotic-wave chromosome functions quite logically. These concern the wave principles of cellular nucleus activity, which, as we argue, explain how the construction of the time-space macrostructure of higher biosystems works along the wave and semantic operational vectors of the genetic apparatus. Such vectors work through mechanisms of a holographic storage of chromosome continuum and through quasi-speech paths of DNA-RNA-structures, which encode the space-time of organisms. The reading or scanning of genome-biocomputer is then executed by means of endogen laser radiation and soliton arousals of gene structures. Genome nonlocality is already included in its holographic information [39,41]. Such sort of information is distributed in the genome as in the hologram and/or quasi-hologram, and - as in a fractal structure - is simultaneous. It can take place, if the genome is interpreted from material positions only. At such level of the genetic information the quantum wave nonlocality does not work yet. If the genetic information is scanned by the wave method, the substance of chromosomes illuminates the semiotic-figurative (sign) wave front sets as the directing vectors (programs) of the morphogenesis. Particularly, this is necessary for maintaining a stable time-space structure of the biosystem. With this purpose, the genome generates stage by stage and layer wise the scheme of potential material frames of an organism some kind of a "theoretical" (wave) model - a plan of potential material organism structures. It is only one of the wave vectors by construction of multivariate frame of the biosystem. In this view, the model of a material-wave organization of biosystems is not complete yet and needs further development.

In relation to the development of these wave notions about genome's semiotic areas of higher organisms, the EPR-mechanism can function, at least, at a level of photon laser and radio wave processes in chromosome and proteins of organisms. The EPR-mechanism, which manages the vital processes, gives totally new potencies to cells and tissues, namely the capacity to actually instantaneously transmit huge information pools between all cells and tissues of the biosystem, for example, through the polarization channel of photons and radio waves. If a such way is possible, it would be the explanation, of why the strategic sign biomolecules - nucleic acids and the protein - have a L-isomeric composition of elements, spiral curling and, accordingly, extremely expressed ability to a dispersion of optical gyration, circle dichroism and birefringence. According to this interpretation, the fact of isomeric quantum nature of bioorganic molecules gains a new quality. The asymmetry of bioorganic molecules atoms and the isomerism, caused by it, means that the biosystem has a possibility of a fast auto scanning of polarization, holographic and other material-wave information on the state of its own metabolism and its own current momentary time-space structure. From this point of view, of unexpected importance for the explanation of prion pathogenesis mechanisms (Creutzfeld-Jacob syndrome, family insomnia, mad cow decease, so called "khourou" illness) gains the ability to birefringence aggregates PrPsc (prion proteins), i.e. to an abnormal modulation of vectors responsible for the polarization of its own informational photon currents through an increasing protein mass of PrPsc in the brain.

The success of the experimental quantum teleportation was achieved, in particular, because wave-guides (light-guides), lasers with ultra-violet pump and polarizers were used to generate photons, to spread them in space and to "program" them. The above mentioned components have formally bioanalogies in form of microtubules of the a cell nucleus and cytoplasm, coherent DNA and chromosome radiation. Simultaneously, the latter are information biopolarizers of their own laser radiation. The proof, that the DNA and chromosomes is a laser active environment, was given in our direct experiments [40].

Let us suppose, that the EPR-factor exists *in vivo* as the controlling factor of a current status of an adult organism from the micro up to macrolevel. How it is implemented in embryogenesis? It could serve as an intermediary for the intracellular and intercellular transmission of wave copies of DNA-RNA in different phases of their polysyllabic operation. The wave memory effects, obtained by us in 1985 and 1991 on the basis of DNA preparations and separately by the Pecora group in USA in 1990, might be a result of the local quantum teleportation, which takes place spontaneously at a laser probing of DNA gels during the spectroscopy by a dynamic laser light distributing method. In this variant of interaction between coherent photons and biostructures, the latter could probably appear as a mesomorphic system of optically active light guides spreading polarized photons in space and interchanging information subsequently between them. In the same series of experiments, another effect with a new type of genetic structures memory was detected on the basis of an Fermi-Pasta-Ulam-phenomenon[27]. It is accompanied by the emergence of isomorphously temporary autocorrelation functions of light distribution during the investigation of preparations of DNA, 50S ribosome subunits E.coli and collagen [27].

If the EPR-factor works in biosystems, it is legitimate to question, why organisms are not restricted to this very efficient form of handling bioinformation and why do they need nervous impulses too, whose velocity of passing (8-10 m/sec) falls far behind the light speed in the DNA quantum biocomputer of living cells? This can be explain by the nature of quantum teleportation itself, since any instantaneous transfers of states, must also be accompanied by corresponding classical transfers at the speed of light or lower. It can also be argued, that higher organisms need the nervous system to slow down information processes, which are too fast and could not be matched by the level of the biosphere evolution. These functions of the nervous system and the genome's quantum nonlocality are thus complementary and must coexist.

References

1. Lukashin A.V. et al, J. Biomol. Struct. and Dyn.,v. 6, N6, pp.1123-1133 (1989).
2. Гаряев П.П., Языковый плюрализм ДНК. (рукопись)
3. Welch G.R., J. Mol. Biol., v.41, N1-4, pp.31-32 (1989).

4. Эйген М., Шустер П. Гиперциклы. Принципы самоорганизации молекул. М. Мир. (1982).

5. Shcherbak V.I., J.Theor.Biol., v.132. pp.121-124. (1988).

6. Katsikas A.A., Nikolis J.S., Nuovo cimento D, v.12, N2, pp.177-195. (1990).

7. Гаряев П.П., Васильев А.А., Березин А.А., Геном как голографический компьютер. Гипотеза. Независ. науч.журн. N1, N1, с.24-43, 49-64 (1991, 1992).

8. Маковский М.М., Лингвистическая генетика. М. Наука. 189с. (1992).

9. Gariaev P.P., Chudin V.I., Komissarov G.G., Berezin A.A., Vasiliev A.A., Proc. SPIE - The Internation. Soc. for Optic. Eng. Optical Memory and Neural Networks. USA. v.1621, pp.280-291. (1991).

10. Solovyov V.V., Korolev S.V., Lim H.A., Int.J.Genome Res., v.1, p.108 (1992).

11. Berthelsen C.L., Glazier J.A., Skolnik M.H., Phys.Rev.A, v.45, N12, pp.8902-8913. (1992).

12. Гаряев П.П., Горелик В.С., Моисеенко В.Н., Попонин В.П., Чудин В.И., Щеглов В.А., Крат. сообщ. по физике Физ. Инст. РАН, N1-2, с.33-36 (1992).

13. Гаряев П.П., Григорьев К.В., Васильев А.А., Попонин В.П., Щеглов В.А., Крат. сообщ. по физике Физ. Инст. РАН, N11-12, с.63-69 (1992).

14. Агальцов А.М., Гаряев П.П., Горелик В.С., Щеглов В.А., Квант. электроника, т.20, N4, с.371-373 (1993).

15. Трубников Б.А., Гаряев П.П., Семиотика ДНК. Российский научный центр. "Курчатовский институт". Препринт. ИАЭ-5690/1. 27с.(1993).

16. Tseng H., Green H., Mol.Biol.and Evol., v.6, pp.460-468 (1989).

17. Федер Е. Фракталы. М. (1991).

18. Jeffrey H., Nucl. Acids Res., v.18, pp.2163-2170 (1990).

19. Соловьев В.В., Королев С.Е., Туманян В.Г., Лим Х.А., ДАН, т.319, N 25, с. 1496-1500. (1991).

20. Королев С.Е., Соловьев В.В., Туманян В.Г. Биофизика, т.37, вып. 5, с. 837-847 (1992).

21. Ратнер В.А., Концепция молекулярно-генетических систем управления. Новосибирск. 118с. (1993).

22. Трубников Б.А., Румынский И.А., ДАН СССР, т.321, N2, 270-272 (1991).

23. Поликарпов А.А. Элементы теоретической социолингвистики. Изд. Моск. Гос. Ун-та (1979).

24. Polikarpov A.A. On the Model of Word Life Cycle. // Contributions to Quantitative Linguistics. - Dordrecht: Kluwer (1993).

25. Поликарпов А.А. Закономерности жизненного цикла слова и эволюции языка. // Русский филологический вестник. М., Московский лицей, № 1 (1994)

26. Петров Н.Б., Маслов М.Ю., Гаряев П.П., Эволюционный анализ рибосомных РНК-текстов. (1999) (рукопись)

27. Гаряев П.П. Волновой геном. М., 279с. (1994).

28. Гаряев П.П., Внучкова В.А., Шелепина Г.А., Комиссаров Г.Г. Вербально-семантические модуляции резонансов Ферми-Паста-Улама как методология вхождения в командно-образный строй генома.// Журнал русской физической мысли. № 1–4, с. 17–28 (1994).

29. Гаряев П.П., Волновой генетический код, "Издатцентр", 107с. (1997).

30. Маковский М.М.. Лингвистическая генетика. М.,1992.

31. Chomsky N. Reflections on Language. N-Y, 1975.

32. Scherbak V. I. // J. Theor. Biol. 1988.V.132.P.121 ?124.

33. Соломоник А. Семиотика и лингвистика. М., 1995. С .345.

34. Einstein A., Podolsky B., Rosen N., Can Quantum-Mechanical Description of Physical Reality be Considered Complete? Phys. Rev. v.47, p.777-780. (1935).

35. Bell J.S., On the Einstein-Podolsky-Rosen Paradox, Physics, v.1, pp.195-200 (1964); Bell J.S., Einstein-Podolsky-Rosen Experiments. Proceed. Symp. on Frontier Problems in High Energy Physics (Pisa), p.33-45 (1976).

36. П.П.Гаряев, М.Р.Гарбер, Е.А.Леонова, Г.Г.Тертышный, К вопросу о центральной догме молекулярной биологии. Сознание и физическая реальность, т.4, №1, с.34-46 (1999).

37. Bouwmeester D., Pan Jian-Wei, Mattle K., Eibl M., Weinfurter H., Zeilinger A., Experimental quantum teleportation. Nature, v.390, p.575-579 (1997).

38. Josephson B.D., Pallikari-Viras F., Biological Utilisation of Quantum NonLocality. Foundation of Physics, v.21, p.197-207, Plenum Press © (1991).

39. Гаряев П.П., Тертышный Г.Г., Готовский Ю.В., Леонова Е.А., Голографическая и квантовая нелокальность генома. V международная конференция "Теор. и клинич. Аспекты применения биорезонансной и мультирезонансной терапии", тезисы и доклады, часть II. "ИМЕДИС", Москва, с.256-272 (1999).

40. Агальцов А.М., Гаряев П.П., Горелик B.C., Рахматуллаев И.А., Щеглов В.А., Двухфотонно-возбуждаемая люминесценция в генетических структурах. Квантовая электроника, v.23, N2, с.181-184 (1996).

41. И.В.Прангишвили, П.П.Гаряев, Г.Г.Тертышный, В.В.Максименко, А.В.Мологин, Е.А.Леонова, Э.Р.Мулдашев, 2000, Спектроскопия радиоволновых излучений локализованных фотонов: выход на квантово-нелокальные биоинформационные процессы. Датчики и Системы, №9(18), с.2-13.

42. Peter P. Gariaev, Boris I. Birshtein, Alexander M. Iarochenko, Peter J. Marcer, George G. Tertishny, Katherine A. Leonova, Uwe Kaempf, The DNA-wave Biocomputer, International Journal of Computing Anticipatory Systems, Edited by Daniel M. Dubois, Published by CHAOS, Liege,Belgium, volume 10, 290-310, ISSN 1373-5411, 1SBN 2-9600262-3-3.

43. Peter Marcer and Edgar Mitchell, What is Consciousness? The Physical Nature of Consciousness, Edited by Philip Van Loocke, John Benjamins Publishing Company, Amsterdam.145-174, in particular 159-160.